



La contribution de l'épigénétique dans l'expression des variants phénotypiques essentiels pour l'interaction : *Schistosoma mansoni* / *Biomphalaria glabrata*

Sara Fneich

► To cite this version:

Sara Fneich. La contribution de l'épigénétique dans l'expression des variants phénotypiques essentiels pour l'interaction : *Schistosoma mansoni* / *Biomphalaria glabrata*. Evolution [q-bio.PE]. Université de Perpignan Via Domitia, 2014. Français. NNT : . tel-01163671

HAL Id: tel-01163671

<https://theses.hal.science/tel-01163671>

Submitted on 15 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par
UNIVERSITE DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale ED 305

Et de l'unité de recherche UMR 5244
Laboratoire Ecologie et Evolution des Interactions 2EI

Discipline : **Biologie**
Spécialité : **Epigénétique**

Présentée par **Sara FNEICH**

La contribution de l'épigénétique dans
l'expression des variants phénotypiques
essentiels pour l'interaction :
Schistosoma mansoni / *Biomphalaria glabrata*

Soutenue le 11 décembre 2014 devant le jury composé de

Mme Christine CHEVILLON, CR1, MiVEGEC UM2
M. Stéphane MAURY, Professeur, LBLGC Orléans
M. Emmanuel ROGER, MCF, CIIL Lille
M. Julio SAEZ VASQUEZ, DR2, LGDP UPVD
M. Christoph GRUNAU, Professeur, 2EI UPVD
Mme Céline COSSEAU, MCF, 2EI UPVD

Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse
Co-directeur de thèse

" Le succès n'est pas la clé du bonheur.

Le bonheur est la clé du succès.

Si vous aimez ce que vous faites vous réussirez. "

Albert Schweitzer

Remerciements

Je souhaite tout d'abord adresser mes remerciements à toutes les personnes qui ont contribué à la réalisation de ce travail ! C'était une très belle aventure de trois ans avec vous...

Merci aux membres du jury qui ont accepté d'évaluer mon travail de thèse. Merci à Christine Chevillon et Stéphane Maury en tant que rapporteurs & à Emmanuel Roger et Julio Saez Vasquez en tant qu'examineurs.

Un merci spécial au directeur du laboratoire. Guillaume, merci de m'avoir fait confiance et m'avoir proposé ce sujet de thèse. J'aimerais également te remercier pour ton accueil au sein du laboratoire « Ecologie et Evolution des Interactions ». Malgré toutes tes occupations, tu étais tout le temps disponible à notre écoute. Merci !

Le plus grand MERCI, je l'adresse à mon directeur de thèse. Christoph, tu es un super encadrant ! Merci pour ta présence aussi bien scientifique qu'humaine. Merci de m'avoir fait confiance et soutenue à tous les moments. Merci pour toutes les discussions et les conseils qui m'ont permis d'évoluer. Tu as toujours trouvé les bons exemples pour me remotiver aux moments où j'étais « épigénétiquement variable », influencée par le stress ! Merci pour tout ce que tu m'as apporté pendant ces trois années ...

Tous mes remerciements s'adressent aussi à ma co-directrice. Céline, merci pour ta présence et ton enthousiasme. Je suis extrêmement sensible à tes qualités humaines d'écoute et de compréhension. Ta capacité à accomplir une multitude de tâches en même temps m'a toujours impressionnée (je ne sais toujours pas comment tu fais !!).

André, j'ai été heureuse de travailler avec vous. Merci pour toutes les discussions et pour les moments passés à la salle du vivant. Un Merci spécial à toutes vos réponses à mes questions infinies...

Pour tous les membres du 2EI, vous étiez ma famille...

Je ne sais pas par qui commencer ! S'il y avait un Top 10, beaucoup parmi vous se bousculeraient pour la première place !! Donc je commence bureau par bureau dès l'entrée jusqu'à notre salle des doctorants...

Benjamin et Jérôme, je sais que mes pâtisseries sont délicieuses ! Mais, je vous rappelle que le sucre n'est pas bon pour la santé !! Benjamin, même si je ne continue pas dans les interactions hôtes / parasites, j'aurai toujours des questions à te poser! Jérôme, je t'assure que j'ai toujours été attentive aux cours de stats ! Merci tous les deux pour toutes les discussions...

Anne M. merci pour tous les moments passés ensemble à la fête de la science, zumba... Diane, ma super voisine du bureau ! Irma, ma mère française ! Et notre super Rémi ... merci pour votre présence aussi bien administrative qu'amicale.

Nathalie ! D'ailleurs, ta formation à Marseille s'est-elle bien passée ? Merci pour tous les moments passés en salle du vivant, les fous rires, les blagues... Bernard, je ne te souhaite que la bonne guérison et merci à Pascal de continuer ce boulot...

Anne R. ! Merci pour tous les moments que tu m'as accordés, pour apprendre les astuces de récupération maximale des parasites mais aussi pour m'écouter râler ;) c'était agréable de te connaître ... Hélène et Gabi, merci pour toutes les discussions, l'accompagnement pendant les weekends et les manips « tardives ! » ...

La plus belle équipe David D, Ricardo et Jeff... Un très grand merci pour touuuuus les moments passés ensemble, à la paillasse et aux pauses. Mes résultats devaient passer par vous avant que je ne retourne à mon bureau. David D., merci de ton soutien scientifique et amical aussi, ça m'a fait plaisir de travailler avec toi également ! Merci pour toutes les discussions, les manips et la relecture du manuscrit... Ricardooooo !!! Merci pour ta présence scientifique et m'écouter râler aussi !... Ne t'inquiètes pas, il y aura toujours des libanais qui passeront par le labo... Jeff notre encyclopédie technique, merci pour toi aussi...

Juliette, c'est vrai les hybrides F2 et F3 sont de vrais hybrides ;) merci à toi pour toutes les discussions intéressantes... Eve, Merci pour les essais infinis sur « Word »... Et pour le

*Super « Journal Club », très bien organisé et très intéressant ! Préparer un papier c'est bien..
Mais être forcé à le préparer... c'est bien AUSSI ;)*

Et mon bureau !! Notre bureau !!

*Marion ma coupine ! Merci pour toutes les soirées manips et rédaction passées ensemble.
Pour tous mes moments Ange/Diable... me manque une Marion ! Bon courage pour toi
aussi ! Rodolfo, franchement même moi, je ne garde plus mon calme dans ce bureau !
Comment tu arrives à rester si calme... Kelly, ohhh Kelly.. Merci pour ta présence à tous les
moments, tu étais tout le temps mon anti-stress ! David R, merci pour ces super 2 années
passées dans la même équipe et pour ta présence à tous les moments (scientifique, découverte
des villes et relecture du manuscrit)... Jérémie, « yeslamo » pour toutes tes connaissances, tu
avais toujours des réponses à nos questions ! Et merci aussi pour ta présence au bureau pour
garder le SILENCE... Benoit, Anaïs et Sylvain bien venue au Club des doctorants.*

BON COURAGE A VOUS TOUS POUR LA SUITE!

*Et ceux que j'ai croisés depuis mon arrivée au labo et ne sont plus là, Julien, Nolwenn et les
2 Julie (s)...*

*Finalement, un très grand Merci à toutes les personnes qui m'ont accompagnées de proche et
de loin. Vous avez fait de ces années une période agréable et inoubliable. Je vous remercie à
tous du fond du cœur ... je ne citerai pas de prénoms parce que je vais sûrement en oublier. A
Tous ceux que j'ai connus de différents laboratoires de l'UPVD et Mes amis Perpignanais
HORS laboratoires. Toute ma famille au LIBAN, toute ma famille en BELGIQUE et Ma
PETITE famille.*

A mon fils Hassouna

إلى حسونة الغالي

Préambule

**La contribution de l'épigénétique dans l'expression des variants phénotypiques
essentiels pour l'interaction :**

Schistosoma mansoni / Biomphalaria glabrata

Cette thèse a été effectuée au sein du laboratoire 2EI à Perpignan dans l'Axe 3: Épigenétique Environnementale et Évolutive des interactions. Comme vous le remarquez dans le titre, le sujet de thèse se situe à l'interface de trois thèmes différents : **Épigénétique, Variants Phénotypiques et Interactions Hôte / Parasite.**

L'épigénétique est aujourd'hui un domaine d'étude émergent qui suscite de plus en plus l'intérêt des évolutionnistes. Les processus évolutifs reposent sur la production des **variants phénotypiques** chez les individus. De nos jours, il est communément accepté que la diversité génétique soit la seule source de la variabilité phénotypique. Cependant, des travaux récents montrent que les modifications épigénétiques contribuent également à la genèse des variants phénotypiques sans engendrer des modifications génétiques. Les **interactions hôte / parasite**, constituent un domaine d'étude très large. Dans ce contexte, il paraît intéressant d'aborder le lien qui peut exister entre l'épigénétique et la variabilité phénotypique rencontré dans le domaine des interactions hôtes / parasites.

L'avantage d'un tel sujet à différents thèmes, c'est que durant ces trois années, j'ai eu l'occasion d'enrichir mes connaissances dans différents domaines. Ainsi, cela m'a permis de participer à différents types de congrès pour présenter mon travail sous différents angles (Epigénétique : SMBE et Epigenetics in Ecology and Evolution ; Interactions : REID et IMH ; Parasitologie : 13^{ième} International Symposium on Schistosomiasis et Workshop on Schistosomiasis).

Au cours de cette thèse, je me suis intéressée surtout à l'étude des modifications épigénétiques chez le parasite *S. mansoni* et son hôte *B. glabrata*.

Je vous souhaite une bonne lecture...

Table des matières :

CHAPITRE I : INTRODUCTION GENERALE.....	1
I.1. L'ADAPTATION.....	3
I.2. ADAPTATION ET VARIABILITE PHENOTYPIQUE.....	3
I.2.1. Origine et héritabilité variants phénotypiques.....	4
I.2.1.1. Les théories de l'évolution adaptative	4
I.2.1.2. Origines des variants phénotypiques	5
I.2.1.3. Exemples d'héritabilité non mendélienne	6
I.2.2. Les deux systèmes d'héritabilité « <i>Dual inheritance system</i> ».....	8
I.3. NOTION D'EPIGENETIQUE	9
I.3.1. Porteurs de l'information épigénétique.....	11
I.3.1.1. La méthylation de l'ADN.....	12
I.3.1.1.1. La 5-méthylcytosine (5mC)	12
I.3.1.1.2. Méthylation de l'ADN chez les vertébrés <i>vs</i> les invertébrés.....	13
I.3.1.1.3. La 5-hydroxyméthylcytosine (5hmC).....	17
I.3.1.2. Les modifications des histones.....	18
I.3.1.3. Les ARN non codants.....	21
I.3.1.4. La localisation nucléaire des loci	21
I.3.2. Etat chromatinien : Euchromatine / Hétérochromatine	22
I.4. MODELE THEORIQUE : GENETIQUE / EPIGENETIQUE / VARIABILITE PHENOTYPIQUE..	23
I.5. INTERACTIONS HOTES / PARASITES : BON MODELE POUR LA BIOLOGIE EVOLUTIVE	25
I.5.1. Modèle d'étude : <i>Biomphalaria glabrata</i> / <i>Schistosoma mansoni</i>	25
I.5.1.1. La bilharziose : Deuxième endémie parasitaire mondiale	25
I.5.1.2. Cycle de vie de <i>S. mansoni</i>	26
I.5.2. La théorie de la Reine Rouge.....	30
I.5.2.1. « <i>Gene for gene</i> ».....	32
I.5.2.2. « <i>Matching phenotype</i> »	32
I.5.3. Polymorphisme de compatibilité.....	34
I.5.3.1. Polymorphisme de compatibilité – traits de vie	36
I.5.3.2. Polymorphisme de compatibilité moléculaire	38

I.5.4. Mécanismes de diversification des éléments clés de l'interaction.....	42
I.5.5. Régulation épigénétique des interactions hôte / parasite	43
I.5.5.1. Epigénétique chez <i>S. mansoni</i> et <i>B. glabrata</i> : Etat des connaissances	44
I.5.5.1.1. Epigénétique chez <i>S. mansoni</i>	45
I.5.5.1.2. Epigénétique chez <i>B. glabrata</i>	47
I.5.5.2. Régulation épigénétique des SmPoMucs chez <i>S. mansoni</i>	48
I.6. OBJECTIFS DE LA THESE	50
CHAPITRE II : IMPLICATION DE L'EPIGENETIQUE DANS LE SUCCES D'INFESTATION CHEZ LE PARASITE SCHISTOSOMA MANSONI	53
II.1. PARTIE 1 : LE ROLE DE L'EPIGENETIQUE DANS LA REGULATION DES SMPOMUCS.....	56
II.1.1. Contexte de la partie 1	56
II.1.2. Publication N°1: <i>Schistosoma mansoni</i> Mucin Gene (<i>SmPoMuc</i>) Expression: Epigenetic Control to Shape Adaptation to a New Host.....	58
II.2. PARTIE 2 : LES BASES EPIGENETIQUES DES VARIANTS PHENOTYPIQUES ADAPTATIFS CHEZ LE PARASITE <i>SCHISTOSOMA MANSONI</i>	64
II.2.1. Contexte de la partie 2	64
II.2.2. Approches méthodologiques.....	66
II.2.2.1. Les croisements et génération des hybrides de <i>S. mansoni</i>	66
II.2.2.2. Le traitement de <i>S. mansoni</i> à la Trichostatin A (TSA)	69
II.2.3. Publication N°2 : Epigenetic origin of adaptive phenotypic variants in the blood-fluke <i>Schistosoma mansoni</i>	72
II.2.4. Résultats complémentaires non publiés	73
II.2.5. Tests statistiques	77
CHAPITRE III : MISE EN EVIDENCE DES PORTEURS DE L'INFORMATION EPIGENETIQUE CHEZ L'HOTE BIOMPHALARIA GLABRATA	79
III.1. OBJECTIF SCIENTIFIQUE DU CHAPITRE III.....	81
III.2. APPROCHE METHODOLOGIQUE.....	82
III.3. PUBLICATION N°3: 5-METHYL-CYTOSINE AND 5-HYDROXY-METHYL-CYTOSINE IN THE GENOME OF <i>BIOMPHALARIA GLABRATA</i> , A SNAIL INTERMEDIATE HOST OF <i>SCHISTOSOMA MANSONI</i>	84

III.4. RESULTATS COMPLEMENTAIRES NON PUBLIES	94
CHAPITRE IV : DISCUSSION	99
IV.1. EPIGENETIQUE ET ADAPTATION CHEZ <i>S. MANSONI</i>	102
IV.2. LES MODIFICATIONS EPIGENETIQUES CHEZ <i>B. GLABRATA</i>	117
CHAPITRE V : CONCLUSIONS ET PERSPECTIVES	121
CHAPITRE VI : REFERENCES BIBLIOGRAPHIQUES	129
CHAPITRE VII : ANNEXE	148

Chapitre I

Introduction générale

Chapitre I : Introduction générale

I.1. L'adaptation

L'environnement est en perpétuelle variation, toutes les espèces sont soumises à des contraintes. Ces contraintes sont générées par d'autres individus de la même ou d'une autre espèce (biotiques), et /ou par des pressions physico-chimiques de l'environnement (abiotiques). En réponse à ces contraintes, le comportement, la physiologie et / ou la morphologie d'un individu sont modifiés afin d'accomplir leurs fonctions (Jablonka *et al.* 1995; Thomas *et al.* 2010).

Par définition, l'adaptation est le changement ou l'ajustement d'un caractère ou d'un comportement d'un individu pour qu'il soit en harmonie avec le milieu environnant. Cette définition pourrait aussi être attribuée à l'acclimatation. Cependant, l'acclimatation est une modification ponctuelle et souvent réversible de l'individu, alors que l'adaptation est un processus évolutif sur plusieurs générations (Cuénot 1925). Le caractère adaptatif permettra à l'individu de survivre dans le nouvel environnement et transmettre cette information à sa descendance (Combes 1988; Thomas *et al.* 2010).

L'adaptation s'accompagne de changements phénotypiques de traits d'histoire de vie. L'individu doit être, en effet, capable d'exprimer différents variants phénotypiques. Parmi ces variants phénotypiques, le phénotype ayant la valeur sélective optimale sera sélectionné et transmis aux générations suivantes (Lachmann & Jablonka 1996; Agrawal 2001). La valeur sélective est définie classiquement par le nombre de descendants viables produits par un individu au cours de sa vie. Elle peut être estimée par le produit de la viabilité (survie), et du nombre de nouveau-nés (fécondité) (Thomas *et al.* 2010).

I.2. Adaptation et variabilité phénotypique

L'adaptation nécessite donc une production des variants phénotypiques. L'expression de ces variants chez un individu est influencée par l'environnement (Via & Lande 1985). Il est communément admis que le phénotype P est le résultat des

interactions génotype G - environnement E ($P=G \times E$) (Comstock & Moll 1963). Un ou certains variants phénotypiques participent à l'expression d'un caractère particulier, sous une condition environnementale particulière. Le résultat immédiat d'expression de ces variants phénotypiques détermine le caractère acquis. Le caractère acquis est considéré comme un de la gamme des phénotypes qui pourraient être exprimés chez un individu suite à des conditions défavorables (Smith 1989).

I.2.1. Origine et héritabilité variants phénotypiques

L'origine et l'héritabilité des variants phénotypiques sont des concepts controversés depuis le XVIII^{ème} siècle. L'héritabilité des caractères soumis à la sélection est à la base de l'adaptation. Cette notion a été abordée depuis l'apparition des deux principales théories de l'évolution adaptative.

I.2.1.1. Les théories de l'évolution adaptative

Jean Baptiste Lamarck (1744-1829) a proposé la théorie de l'évolution : « Tous les individus d'une population pourraient acquérir une adaptation suite à un changement environnemental. Leurs descendants seront de plus en plus adaptés à cet environnement » (Lamarck 1815). Pour lui, l'héritabilité d'un caractère est directement influencée par les facteurs environnementaux. Contrairement à ce qui est couramment admis, Lamarck n'a pas inventé le concept de l'héritabilité des caractères acquis, mais ceci faisait partie de ses pensées (Jablonka & Lamb 1999).

Cinquante ans après Lamarck, Charles Darwin (1809-1882) propose sa théorie de l'évolution basée sur la sélection naturelle. D'après Darwin : « Les individus sont différents entre eux, et l'individu ayant des caractéristiques appropriées à l'adaptation, survivra et se reproduira. Il transmettra par la suite ces caractères à la descendance » (Darwin 1859, 1868).

Avant l'apparition de la génétique, Darwin avait proposé une théorie sur l'hérédité. Sa théorie est basée sur l'hypothèse de la « pangenèse ». Pour lui, les différentes parties du corps émettent des particules, qu'il nomme « gemmules », tout au long de la vie de l'organisme. Ces gemmules s'agrègent dans les organes génitaux pour

former les cellules germinales. Le matériel de l'hérédité sera transmis par les gemmules, des parents aux descendants (Darwin 1868).

Les règles de l'hérédité génétique ont été découvertes par Gregor Mendel (1822-1884). Mendel était le premier à réaliser des expériences convaincantes dans le sens où certains caractères héréditaires se maintiennent à la descendance. Ses travaux ont été redécouverts et publiés en 1900 (De Vries 1904).

A la lumière de la génétique, la théorie darwinienne a été réinterprétée par le néo-darwinisme. Pour les néo-darwiniens, les individus portent des bagages génétiques différents. Le matériel héréditaire subit des modifications par des mutations. Ces mutations sont à l'origine des variations phénotypiques (Jablonka & Lamb 1999; Egloff & Fehr 2011).

I.2.1.2. Origines des variants phénotypiques

Il était couramment accepté que la diversité génétique soit la seule source de variabilité phénotypique. Cependant, certains cas de transmissions non mendéliennes sont observés (Frésard *et al.* 2013; Norouzitallab *et al.* 2014). Plus particulièrement, des caractères acquis montrent une héritabilité indépendante des mécanismes génétiques, allant ainsi à l'encontre de la théorie néo-darwinienne. En effet, ces résultats suggèrent que les variations phénotypiques héréditaires ne seraient pas toujours générées par des mutations génétiques (des exemples sont fournis plus tard).

Ceci a motivé la curiosité scientifique vers *au-delà des bagages héréditaires*.

En 1963, Ernst Mayr était le premier à populariser pour la première fois l'expression « hérédité des caractères acquis ». Pour lui, « c'est un type d'héritabilité ou le matériel héréditaire n'est pas constant d'une génération à l'autre, mais pourrait être modifié par les effets de l'environnement, l'utilisation et non utilisation ou d'autres facteurs » (Mayr 1963; Egloff & Fehr 2011).

Les investigations sur les différenciations cellulaires et le développement embryonnaire sont devenues de plus en plus intenses à partir de 1960. En 1961,

l'inactivation du chromosome X ainsi que sa transmission à l'état inactif au cours de plusieurs divisions cellulaires a été démontrée (Lyon 1961). C'était une évidence de la transmission d'un phénotype cellulaire acquis durant le développement, au cours de divisions cellulaires.

Depuis cette période, d'autres études ont montré des cas d'héritabilité phénotypiques qui ne respectaient pas les lois mendéliennes.

1.2.1.3. Exemples d'héritabilité non mendélienne

Des études expérimentales montrent que l'introduction de prédateurs dans l'environnement pourrait provoquer une évolution rapide des mécanismes de défense transmis à la descendance, aussi bien chez des plantes que chez des métazoaires.

Chez les plantes, de tels cas ont été observés par exemple chez *Mimulus guttatus* (la mimule), qui est une espèce de plante sauvage à fleur jaune des zones humides. Lorsque la mimule est attaquée par des herbivores, les dégâts provoqués dans ses tissus conduisent à l'expression d'un phénotype à « trichome plus dense » (figure 1). Les trichomes chez les plantes sont des excroissances épidermiques ressemblant aux poils, ils servent à la défense de la plante contre les herbivores. Ce phénotype est transmis à plusieurs générations même quand elles ne sont pas confrontées aux prédateurs, laissant la séquence d'ADN intacte (Scoville *et al.* 2011).



Figure 1 : Les différents phénotypes de *Mimulus guttatus* en fonction du stress environnemental. (A) La forme typique des mimules, et (B) la forme stressée d'une mimule d'après (Scoville *et al.* 2011).

Parmi les métazoaires, un exemple est observé aussi chez le petit crustacé d'eau douce *Daphnia cucullata* (la daphnie). Dans un environnement serein, la daphnie montre une tête arrondie. Cependant en présence de prédateurs, elle réagit à un signal chimique en s'allongeant et développant un appendice en forme de casque pointu (figure 2), la rendant moins accessible pour ces prédateurs. Ce phénotype est retrouvé dans les générations issues de ces crustacés même en l'absence des prédateurs (Agrawal *et al.* 1999).



Figure 2 : Les différents phénotypes de *Daphnia cucullata* en fonction du stress environnemental. Microscopie électronique d'une daphnie typique (à droite) et d'une daphnie induite (à gauche) (Agrawal *et al.* 1999).

Il est important de noter que chez les animaux aussi bien que chez les plantes, les descendants de la population gardent à leur tour le phénotype « défense contre les prédateurs » sur certaines générations. Egalement, d'autres exemples d'héritabilité des caractères acquis sont cités aussi, tels que les effets parentaux sur la descendance. Les effets parentaux sont définis par les effets que les parents apportent sur le phénotype de leur descendance indépendamment du génotype (Danchin *et al.* 2011). Ceci a été montré chez les mammifères par l'influence de la nutrition et des soins maternels sur le comportement des descendants aux générations suivantes (Champagne 2008; Burdge & Lillycrop 2010; Dias & Ressler 2014).

De nos jours, il n'existe toujours pas d'idées tranchées sur l'origine de l'héritabilité de la variabilité phénotypique. Le caractère acquis pourrait être héritable ou pas en fonction son importance à la survie et à la reproduction des individus.

Certaines variations phénotypiques sont transmises telles que les exemples cités précédemment (Bossdorf *et al.* 2008; Herrera & Bazaga 2010). Cependant d'autres variations ne sont pas sélectionnées et transmises aux générations suivantes, telles que le bronzage de la peau et la musculature du cycliste. Clairement ces phénotypes ne semblent pas être indispensables à la survie de l'individu ou bien à ses descendants.

1.2.2. Les deux systèmes d'héritabilité « *Dual inheritance system* »

Il a été longtemps admis que l'information génétique codée par la séquence de l'ADN est responsable de la diversité et la variabilité des espèces. Cependant, il est récurrent que la diversité génétique reste insuffisante pour expliquer toutes les diversités entre les espèces, les individus d'une même espèce et au sein de l'individu lui-même possédant différents organes avec un bagage génétique identique dans toutes ses cellules somatiques.

Avec le développement de la biologie moléculaire, des mécanismes visant à expliquer l'hérédité non génétique ont émergé. La modification chimique de la séquence de l'ADN par méthylation des cytosines, en est l'exemple le mieux documenté. La méthylation de l'ADN a permis de montrer à la fois l'hérédité non génétique et la modification de l'expression des gènes (Sager & Kitchin 1975). Cette hérédité est décrite à l'heure actuelle comme « épigénétique », terme inventé par Conrad Waddington (1905-1975) en 1942 (voir I.3. Notion d'épigénétique) (Waddington 1942; Jablonka *et al.* 1992).

La vision de l'hérédité stricte aux gènes était prépondérante jusqu'à l'émergence de la théorie du « *Dual inheritance system* », par John Maynard Smith en 1990. Pour lui, « le caractère acquis c'est que des facteurs environnementaux affectent directement des protéines ou autres structures sans impliquer les acides nucléiques, et cette information sera transmise d'une génération à l'autre, mais la stabilité de ces états

transmis reste faible pour avoir une importance évolutive » (Maynard Smith 1990). Cette théorie est devenue répandue à nos jours, évoquant l'existence de deux systèmes d'héritabilité (Maynard Smith 1990; Danchin *et al.* 2011) :

- 1- L'héritabilité de la séquence de l'ADN : hérabilité génétique.
- 2- L'héritabilité de l'état des gènes incluant l'activation et la différenciation au cours du développement : hérabilité épigénétique.

A l'heure actuelle, l'hérabilité épigénétique est démontrée chez plus d'une centaine d'espèces incluant les bactéries, les protistes, les champignons, les plantes et les animaux. La plupart de ces caractères restent stables au cours des générations. Certains de ces changements épigénétiques, mais pas tous, sont associés à des phénotypes hérables et adaptatifs (Jablonka & Lamb 1995). Ces variations épigénétiques peuvent conduire à l'hérabilité de certains caractères acquis, et peuvent ainsi être considérées comme la base de certains changements évolutifs (Jablonka *et al.* 1992).

En conclusion, non seulement les mécanismes génétiques, mais aussi d'autres mécanismes tels que les mécanismes épigénétiques peuvent être une source des variants phénotypiques.

I.3. Notion d'épigénétique

« Entre le génotype et le phénotype existe des mécanismes complexes de développement », disait Conrad Waddington (1905-1975) en 1942 (Waddington 1942). Waddington, généticien et embryologiste est à l'origine du terme « épigénétique ». Il l'avait proposé pour nommer la discipline qui étudie l'interaction entre les gènes et leur environnement faisant apparaître le phénotype (figure 3).

D'après ses expériences sur *Drosophila melanogaster*, il montra que des stress environnementaux, plus particulièrement le choc thermique, induisent des variations phénotypiques et le caractère acquis est transmis à certains descendants (Waddington 1942). Waddington a appelé ce phénomène « assimilation génétique » (Waddington 1953). Le phénotype apparaît initialement en réponse à un

environnement différent ; la sélection le convertit en caractère génétique permanent. La théorie de Waddington représente une réapparition de la théorie lamarckienne sur l'héritabilité des caractères acquis tout en restant compatible avec l'implication des mécanismes génétiques et la théorie néo-darwinienne (Thomas *et al.* 2010).

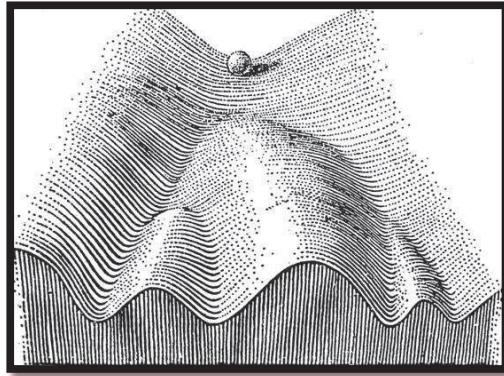


Figure 3: Le paysage épigénétique de Waddington. Les différents chemins parcourus par la bille, vers la gauche, la droite, le haut ou le bas, correspondent à l'histoire du développement d'un embryon. Chacun de ces chemins correspond à une partie développementale de cet embryon (Waddington 1957).

A cette époque, les bases moléculaires de l'épigénétique étaient inconnues. C'est à la fin du XX^{ème} siècle que le terme épigénétique s'est répandu notamment dans les sciences médicales où l'on commence à développer les notions d'épiallèles et d'épimutations (définition plus tard).

Actuellement, l'épigénétique est définie par l'étude des changements héréditaires et réversibles dans la régulation de l'expression des gènes sans faire appel à des modifications de la séquence de l'ADN (Richards 2006; Bird 2007). C'est l'ensemble des processus qui expliquent la plasticité développementale aussi bien chez les procaryotes que chez les eucaryotes (Jablonka & Raz 2009). L'exemple le plus classique c'est le développement et la différenciation des cellules de l'organisme. Toutes nos cellules possèdent le même bagage génétique, pourtant elles se développent et se différencient de différentes manières, établissant des organes différents au niveau morphologique aussi bien que fonctionnel. L'expression différentielle des gènes dépend de différents facteurs, parmi lesquels leur état chromatinien. L'épiallèle est un allèle qui peut exister sous différents états

épigénétiques, résultant différents phénotypes (Rakyan *et al.* 2002). La séquence nucléotidique des différents épiallèles est invariable. Les épimutations consistent en des modifications des états épigénétiques des allèles, générant donc les épiallèles (Horsthemke 2006).

La chromatine est la structure nucléo-protéique sous laquelle se présente l'ADN. Elle a été mise en évidence par Walther Flemming (1843-1905) en 1882 (Flemming 1882). Le nucléosome constitue l'unité de base de la chromatine, qui à son tour est composé d'un octamère d'histones (2 histones H2A, 2 histones H2B, 2 histones H3 et 2 histones H4). L'ADN s'enroule autour des nucléosomes sur une longueur de 147 pb environ, pour former le « collier de perles ». Les nucléosomes sont reliés entre eux par une séquence d'ADN sur laquelle se fixe l'histone H1 dite « protéine de liaison » (Olins & Olins 1974).

Le nucléosome constitue une barrière physique et chimique qui contrôle l'accessibilité de différents facteurs à la séquence d'ADN. Toutes les fonctions de l'ADN : transcription, réplication et réparation sont déterminées par la dynamique des nucléosomes. Ces derniers constituent donc le premier niveau de compaction de l'ADN, par conséquent le degré de compaction de l'ADN joue un rôle primordial dans la régulation de l'expression génique. Cette compaction est surtout contrôlée par les porteurs de l'information épigénétique (Strahl & Allis 2000; Luger 2006).

I.3.1. Porteurs de l'information épigénétique

Quatre supports de l'information épigénétique sont décrits jusqu'à présent, intervenant surtout dans la régulation de la structure chromatinienne et les modifications chimiques de l'ADN. Les deux bases moléculaires les plus étudiées sont : la modification covalente de l'ADN via la méthylation de l'ADN et les modifications des histones incluant les modifications post-traductionnelles et les variants d'histones. Les autres bases moléculaires correspondent aux ARN non codants et la localisation nucléaire des loci (figure 4).

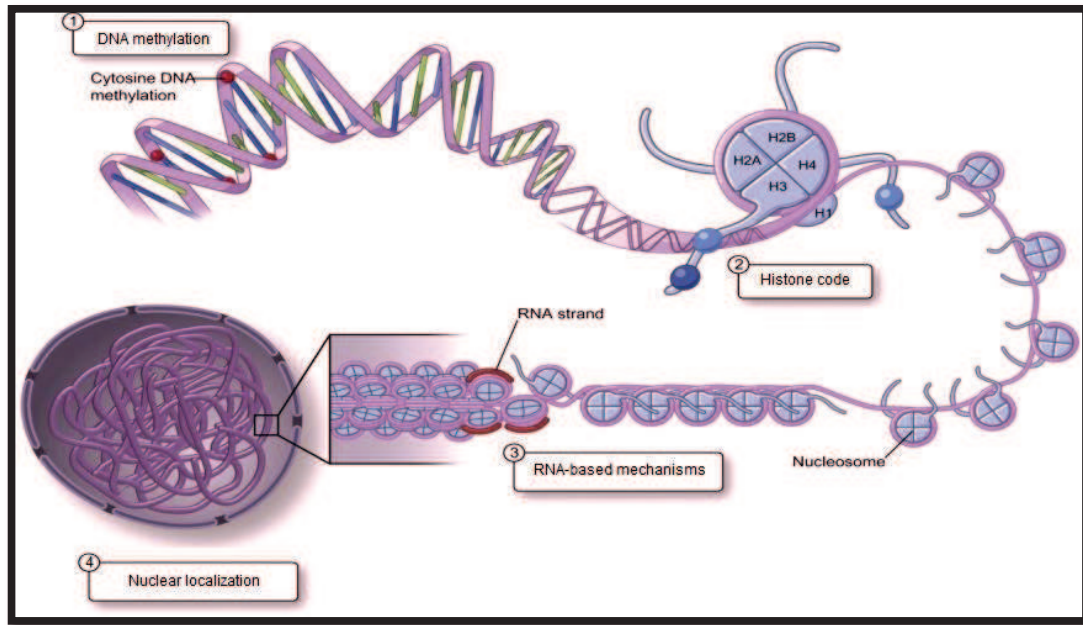


Figure 4 : Les porteurs de l'information épigénétique. La méthylation de l'ADN, les modifications des histones, les ARN non codants et la localisation nucléaire, figure modifiée à partir de (Lovrecic *et al.* 2013).

I.3.1.1. La méthylation de l'ADN

I.3.1.1.1. La 5-méthylcytosine (5mC)

La méthylation de l'ADN est la marque épigénétique la plus étudiée, un mécanisme fondamental présent dans les génomes des animaux, végétaux, champignons et des bactéries (Heithoff *et al.* 1999). C'est une modification chimique qui consiste en l'ajout d'un groupement méthyle (CH_3) sur le carbone en position 5 d'une cytosine à la place d'un atome d'hydrogène pour former la 5-méthylcytosine (5mC), considérée comme la cinquième base de l'ADN (figure 5). La 5-méthylcytosine a été mise en évidence dans les années 1950 chez les animaux et chez au moins une plante supérieure (Wyatt 1951).

Cette modification ne s'effectue pas d'une façon uniforme sur le génome, elle est surtout dépendante de la fonction qu'elle va accomplir (Suzuki & Bird 2008). La méthylation de l'ADN intervient surtout dans la régulation de la transcription des gènes, associée au développement, à la différenciation et la viabilité, à la stabilité du génome en inhibant les éléments transposables et à la régulation différentielle des

allèles connue par l’empreinte parentale (Li *et al.* 1992; Zilberman *et al.* 2006; Lister & Ecker 2009).

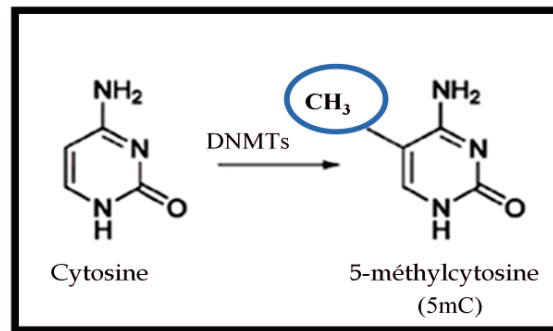


Figure 5 : La méthylation de la cytosine. 5mC est formée par l’addition d’un groupement méthyl à la cytosine via l’action des DNMTs.

I.3.1.1.2. Méthylation de l’ADN chez les vertébrés *vs* les invertébrés

Chez les vertébrés, la méthylation s’effectue surtout sur les cytosines des dinucléotides CpG d’une façon symétrique. Environ 70% des CpG sont méthylés dans les génomes des mammifères (Strichman-Almashanu *et al.* 2002; Hallgrímsson & Hall 2011). Les CpG non méthylés quant à eux, sont regroupés pour former les « îlots CpG » principalement au niveau des régions 5’ des gènes (Bird 1985).

Bien que la méthylation de l’ADN soit spécifique aux dinucléotides CpG chez les vertébrés ; chez les invertébrés, elle n’est pas limitée aux CpG, elle s’effectue également au niveau des autres dinucléotides CpA, CpT et CpG (Salzberg *et al.* 2004).

La réaction de la méthylation est assurée par des ADN méthyltransférases DNMTs. Chez les mammifères, trois classes de DNMTs aux fonctions différentes ont été caractérisées jusqu’à présent. La DNMT1 intervient dans la maintenance de la méthylation de l’ADN, méthylant l’ADN hémi-méthylé (Cheng & Blumenthal 2008). La DNMT2 possède une activité de méthylation des ARNt (Goll *et al.* 2006). Elle possède par ailleurs de fortes similarités de séquence avec la DNMT1 au niveau d’un domaine catalytique. Cependant son activité d’ADN méthyltransférase n’a encore jamais été montré sans ambiguïté et son rôle reste énigmatique (Goll *et al.* 2006).

Finalement, les DNMT3 sont responsables de la méthylation « *de novo* », et sont essentielles au développement et à la viabilité des mammifères (Okano *et al.* 1999).

D'une manière générale, le rôle de la méthylation de l'ADN a été bien caractérisé chez les vertébrés. Il a été démontré que cette modification a lieu essentiellement dans les régions inter-géniques, au niveau des promoteurs, dans les éléments transposables et les séquences répétées. La méthylation de l'ADN est impliquée essentiellement dans la régulation de l'expression des gènes chez les vertébrés et dans la défense des génomes contre les éléments mobiles. En effet, la méthylation des éléments transposables les empêchent de s'exprimer et de se déplacer dans le génome (Li *et al.* 1992; Pedersen *et al.* 1999; Elango 2008; Zemach *et al.* 2010).

Chez les invertébrés, la méthylation de l'ADN montre des degrés d'importance variables en fonction des espèces. D'ailleurs, le taux de méthylation de l'ADN est extrêmement variable au sein des invertébrés. L'ADN peut soit être peu ou pas méthylé comme chez *Drosophila melanogaster* (Urieli-Shoval *et al.* 1982), *Tribolium castaneum* (Zemach *et al.* 2010) et *Caenorabditis elegans* (Simpson *et al.* 1986), alors qu'elle est très présente chez d'autres espèces comme chez *Apis mellifera*, *Crassostrea gigas* et *Acyrtosiphon pisum* (Gavery & Roberts 2010; Lyko *et al.* 2010; Walsh *et al.* 2010).

Bien que le génome d'*A. mellifera* contient les gènes DNMT1, DNMT2 et DNMT3 (Kucharski *et al.* 2008), celui de *C. gigas* par exemple, a DNMT1 et DNMT3, mais pas DNMT2 (Gavery & Roberts 2010). D'autres génomes ne codent que pour DNMT1 et DNMT2 comme celui de *Bombyx mori* (Mita *et al.* 2004). Nous retrouvons également des génomes avec uniquement DNMT2, comme ceux de *Drosophila pseudoobscura* et *Anopheles gambiae* (Marhold *et al.* 2004), ils sont qualifiés « *DNMT2-only* » organisms (Krauss & Reuter 2010). D'ailleurs, des études effectuées chez *Drosophila* ont montré que l'inhibition de la DNMT2 n'affecte pas le développement embryonnaire et la viabilité des drosophiles, sachant que la drosophile fait partie des « *DNMT2-only* » organisms (Kunert *et al.* 2003). Dans d'autres génomes, comme celui de *C. elegans* (Gutierrez & Sommer 2004), les DNMTs sont complètement absentes.

Contrairement aux vertébrés, la méthylation de l'ADN chez les invertébrés s'effectue plutôt dans les régions intra-géniques et particulièrement au niveau des unités transcriptionnelles (Suzuki *et al.* 2007). En ce qui concerne les éléments transposables et les séquences répétées, leur méthylation n'est pas consensuelle comme chez les vertébrés. Chez *B. mori* et *C. intestinalis*, la 5mC est faible voire absente au niveau de la plupart des transposons (Zemach *et al.* 2010).

Ces observations permettent de suggérer que le faible taux de méthylation pourrait faciliter la régulation de la transcription de certains gènes (Roberts & Gavery 2011). Cependant, son rôle reste moins clair chez les invertébrés que chez les vertébrés.

La méthylation des gènes chez les invertébrés n'est pas uniforme. La plupart des espèces invertébrés possède un génome méthylé de type « mosaïque » caractérisé par la présence de deux lots de gènes : les gènes fortement méthylés qui sont souvent liés à des gènes de ménage et les gènes faiblement méthylés souvent liés à des gènes impliqués dans les réponses immunitaires (Suzuki *et al.* 2007; Nanty *et al.* 2011) (Lyko *et al.* 2010; Walsh *et al.* 2010). Ainsi, le taux de méthylation chez les invertébrés est relativement plus faible par rapport à celui des vertébrés (tableau 1) (Tweedie *et al.* 1997).

Les cytosines méthylées sont des bases hypermutables, et sont donc spontanément transformées en thymine par désamination, ce qui conduit à une perte progressive des sites CpG dans les génomes (Coulondre *et al.* 1978; Fryxell & Zuckerkandl 2000). Lorsque les sites CpG sont sous-représentés sur une séquence déterminée (le ratio : CpG/GpC <1), ceci montre que cette séquence est méthylée ou bien a été méthylée chez les ancêtres de l'espèce en question (Antequera & Bird 1993). Par conséquent, les régions méthylées de l'ADN possèdent des ratios CpG/GpC plus faibles que les ratios des régions non méthylées. Ce ratio CpG/GpC peut être utilisé pour prédire la méthylation à partir de données de transcriptomiques. Il est couramment calculé sur les transcrits issus d'approche RNA-Seq chez les invertébrés, chez qui la méthylation a lieu préférentiellement dans les régions codantes. Une bonne corrélation entre les données expérimentales et ces prédictions a été montrée sur plusieurs modèles, notamment chez le mollusque *C. gigas* (Gavery & Roberts 2013). En étudiant la

méthylation de l'ADN à l'échelle du génome entier chez ce mollusque, ils ont démontré que les gènes méthylés (dans leurs régions intra-géniques) sont associés à des états transcriptionnels élevés (Gavery & Roberts 2013).

Par ailleurs, la distribution de ces ratios CpG/GpC sur les transcriptomes de différentes espèces a fait également l'objet d'études intéressantes donnant lieu à des différences de profils entre les vertébrés et les invertébrés.

	Invertébrés	Vertébrés
Méthylation de l'ADN	Non ubiquitaires (incluant les cnidaires, mollusques et échinodermes)	Tous (incluant les poissons, amphibiens, reptiles, oiseaux et mammifères)
Le génome	Méthylés de façon mosaïque (une fraction méthylée et une fraction non méthylée)	Globalement méthylés (la majorité de types cellulaires)
Localisation principale de la méthylation	Régions intra-géniques et unités transcriptionnelles	Régions inter-géniques, promoteurs
Les éléments répétés	Différences entre espèces et entre types de séquences répétées	Tous méthylés
Le ratio CpG/GpC	Bimodale/unimodale suivant les espèces	unimodale
CpG contenant 5mC	10 à 40% de la fraction méthylée	60 à 90%

Tableau 1 : Comparaison de la méthylation des génomes entre invertébrés et vertébrés. Les principales différences de la méthylation de l'ADN entre les invertébrés et les vertébrés (Tweedie *et al.* 1997; Suzuki *et al.* 2007; Elango 2008; Zemach *et al.* 2010).

Chez les vertébrés, la distribution des ratios CpG est généralement unimodale au niveau des régions intra-géniques (Elango 2008). En revanche, chez les invertébrés, la distribution des ratios CpG/GpC peut être unimodale ou bimodale en fonction des espèces. La distribution bimodale est observée chez plusieurs espèces d'invertébrés, dont le mollusque *C. gigas* (Gavery & Roberts 2010). Cette bimodalité est traduite par la présence de deux lots de gènes dans le génome : (i) des gènes fortement méthylés, présentant un faible ratio CpG/GpC, et (ii) des gènes faiblement méthylés présentant

un ratio CpG/GpC qui est élevé (Suzuki *et al.* 2007; Feng *et al.* 2010; Zemach *et al.* 2010).

Ces différences remarquables de la méthylation de l'ADN entre invertébrés et vertébrés, ainsi que toutes les fonctions assurées par cette modification chimique, rendent ce processus essentiel dans la régulation de l'expression des gènes et les phénotypes associés.

1.3.1.1.3. La 5-hydroxyméthylcytosine (5hmC)

La 5 méthylcytosine (5mC) est une base dynamique, qui pourrait être convertie en une autre forme de la cytosine. La 5-hydroxyméthylcytosine (5hmC), forme modifiée de la 5mC, a été décrite dans l'ADN des mammifères dans les années 1970 (Penn *et al.* 1972). Les enzymes Ten-Eleven Translocation (TETs) catalysent la conversion de 5mC en 5hmC en incorporant un groupement hydroxyle (figure 6) (Tahiliani *et al.* 2009; Ito *et al.* 2010). Bien que sa fonction ne soit pas encore claire, la 5hmC est considérée récemment comme la sixième base de l'ADN. Les études récentes montrent que cette base est associée à des unités transcriptionnelles et par conséquent, serait impliquée dans la régulation de la méthylation de l'ADN dans ces régions.

Il a été montré que l'interaction entre l'ADN méthyltransférase DNMT1 et la 5hmC est faible. Cette observation permet de suggérer que la 5hmC empêche l'activité de méthylation de la DNMT1 sur sa cytosine cible. Ceci attribuerait à la 5hmC un rôle potentiel dans la déméthylation de l'ADN (Valinluck & Sowers 2007).

Chez les mammifères, les études montrent que le taux de 5hmC est variable et spécifique en fonction des tissus. La 5hmC est très élevé au niveau des cellules embryonnaires souches et du système nerveux central, notamment au niveau des neurones de Purkinje (Kriaucionis & Heintz 2009; Tahiliani *et al.* 2009). Ces observations suggèrent que la 5hmC intervient dans les stades précoces du développement embryonnaire et de différenciation cellulaire (Ficz *et al.* 2011; Tan & Shi 2012).

Chez les invertébrés, la présence et la fonction de la 5-hydroxyméthylcytosine (5hmC) sont en cours d'étude. Une famille d'enzymes TETs, catalysant la conversion de 5mC en 5hmC, est présente dans la majorité des génomes d'invertébrés séquencés (Pastor *et al.* 2013).

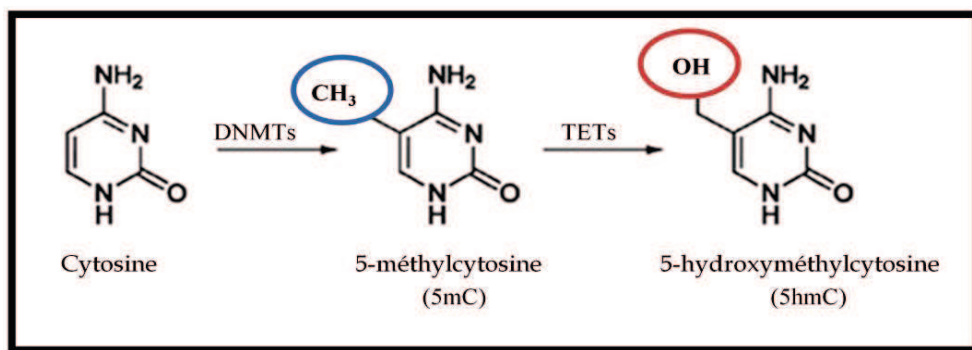


Figure 6 : Différentes formes de la cytosine. La 5mC est formée par l'addition d'un groupement méthyl à la cytosine via l'action des DNMTs. La 5hmC est formée par les TETs protéines, en transférant un groupement hydroxyle à la 5mC.

Une étude très récente a révélé la présence de la 5hmC dans le génome de l'abeille *A. mellifera*, durant la phase de développement. L'unique TET codée par le génome de l'abeille possède la capacité d'oxyder la 5mC en 5hmC (Wojciechowski *et al.* 2014).

Il est évident que la 5hmC et sa machinerie sont présentes chez certains invertébrés. Cependant, comme chez les mammifères, le rôle de cette base reste encore à éclaircir.

I.3.1.2. Les modifications des histones

Les histones, protéines constituant le nucléosome, sont soumises à des modifications post-traductionnelles au niveau des résidus de leurs queues N-terminal sortantes. L'acétylation des lysines, la méthylation des lysines et des arginines (mono, di ou tri), la phosphorylation des sérines et des thréonines, l'ubiquitylation, la sumoylation, l'ADP ribosylation, et la deimination sont les différentes classes de modifications des histones identifiées jusqu'à présent (figure 7) (Kouzarides 2007).

Ces modifications peuvent jouer un rôle « d'activation » ou bien « de répression » en fonction de (i) leur localisation sur un acide aminé particulier et de (ii) l'emplacement des histones mêmes sur la séquence d'ADN.

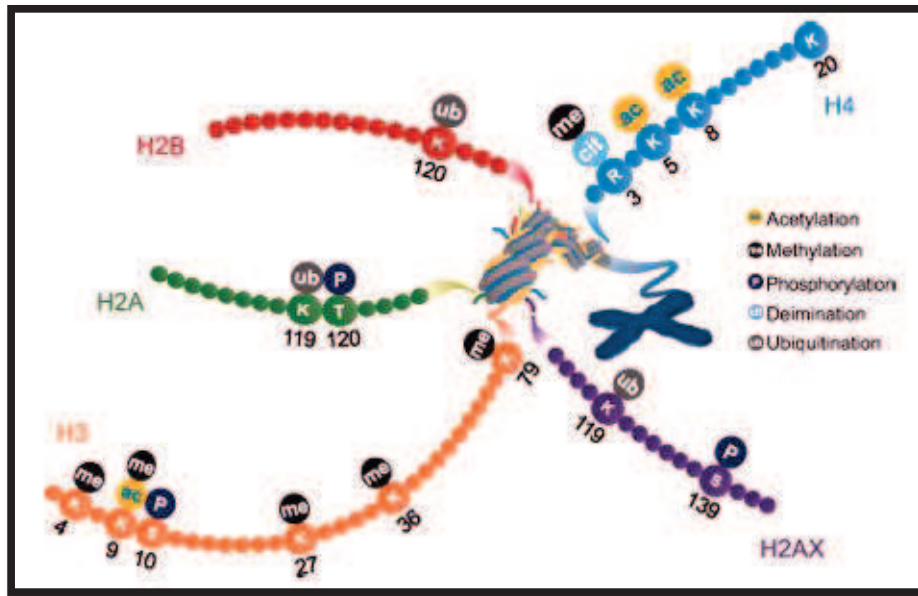


Figure 7 : Représentation schématique des modifications d'histones les plus courantes. Les différentes modifications chimiques des queues des histones canoniques (H2A, H2B, H3 et H4) et non canoniques H2A.X (issu d'Abcam).

(i) L'acétylation de la lysine K9 de l'histone 3 (H3K9Ac) est une marque associée plutôt au point +1 de la transcription *Transcription Start Site* (TSS) et aux régions promotrices du gène jouant un rôle dans l'activation de la transcription (Wang *et al.* 2008). Cependant, elle devient une marque fortement répressive de la transcription du gène, lorsque la lysine k9 est tri-méthylée (H3K9me3). H3K9me3 est associée aux gènes silencieux, localisée dans une région de 10 kb avoisinant le TSS (Berger 2007; Wang *et al.* 2008).

(ii) En termes d'exemple, La méthylation de la lysine 36 de l'histone 3 (H3K36) joue deux rôles différents en fonction de sa localisation dans les gènes. Dans la région codante du gène, la méthylation de H3K36 « active » l'expression, tandis qu'elle « l'inhibe » une fois localisée dans le promoteur (Li *et al.* 2007).

D'autres exemples de corrélation entre les modifications des histones et la transcription des gènes sont illustrés dans la figure 8.

Ces modifications chimiques sont catalysées par différentes enzymes relativement conservées chez les eucaryotes, telles que les histones méthyltransférases (HMTs), les

histones acétyltransferases (HATs) et les histones désacétylases (HDACs) (Zhang & Reinberg 2001; Peserico & Simone 2010).

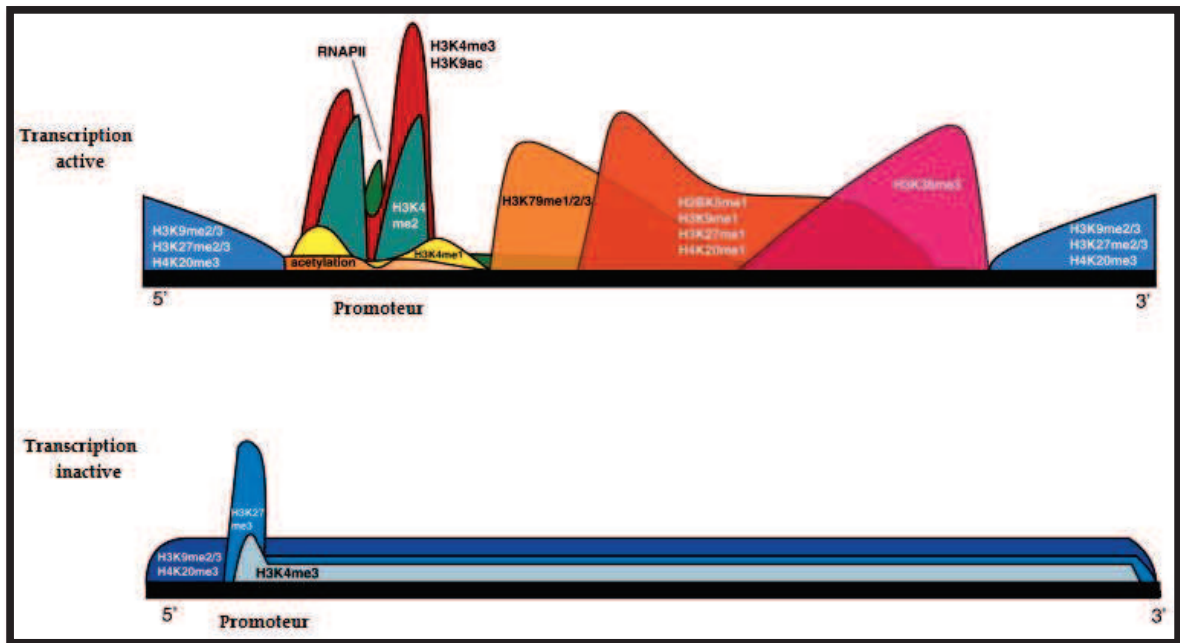


Figure 8 : La distribution des modifications des histones à l'échelle du génome, en corrélation avec la transcription des gènes. Cette figure représente la répartition des modifications des histones sur des gènes quelconques des eucaryotes par rapport au promoteur, d'après (Barth & Imhof 2010).

Les différentes combinaisons de ces modifications forment le « code des histones », qui à son tour, déterminera l'état chromatinien autour d'un locus ou d'un gène (Wang *et al.* 2009). Un locus peut avoir un état chromatinien allant de « ouvert et permissif à la transcription » jusqu'à « fermé et non permissif à la transcription ». Par conséquent, l'état chromatinien d'un locus module l'accessibilité des facteurs de transcription et d'autres protéines régulatrices à la séquence de l'ADN (Narlikar *et al.* 2002).

En plus des modifications post-traductionnelles que peuvent subir les histones, ces derniers peuvent être remplacés par des variants d'histones non canoniques. Les variants d'histones s'incorporent dans le nucléosome à la place d'une histone canonique. Les histones H3 et H2A possèdent des variants d'histones très conservés chez les eucaryotes (Redon *et al.* 2002). Par exemple, H2A.X est un variant d'histone H2A possédant une région C-terminale plus longue que celle de H2A (figure 7 et 8).

H2A.X intervient principalement dans la réparation des cassures de l'ADN double brin (Srivastava *et al.* 2009). Ces variants d'histones assurent au nucléosome une nouvelle organisation accompagnée de nouvelles caractéristiques fonctionnelles (Mizuguchi *et al.* 2004).

I.3.1.3. Les ARN non codants

Les ARN non codants constituent le troisième support de l'information épigénétique. Les études des ARN non codants deviennent de plus en plus massives à nos jours. Ces ARN ne sont pas traduits en protéines, mais la majorité des études montrent qu'ils jouent un rôle essentiel dans la régulation de l'expression des gènes (Mattick & Makunin 2006). Parmi les ARN non codants décrits à l'heure actuelle, nous distinguons les siRNA (small interfering RNA), les miRNA (micro RNA), les piRNA (PIWI-interacting), les lncRNA (long non coding ARN) (Costa 2008), et finalement les snoRNA (small nucleolar RNA) (Bachellerie *et al.* 2002).

Ces différents ARN non codants peuvent diriger les mécanismes de méthylation de l'ADN et les modifications des histones (Bernstein & Allis 2005; Costa 2008; Groszmann *et al.* 2013). D'autres fonctions très importantes leur sont associées, telles que l'inactivation du chromosome X, l'activation /inactivation des éléments transposables et les paramutations (Costa 2005; Greaves *et al.* 2012). Il est important de noter que leurs effets pourraient être héréditaires au cours des générations (Gapp *et al.* 2014) et ils interviennent dans l'hérédité des phénotypes acquis aussi bien chez les vertébrés (Rassoulzadegan *et al.* 2006) que chez les invertébrés (Gu *et al.* 2012).

I.3.1.4. La localisation nucléaire des loci

Dans le noyau des eucaryotes, les chromosomes ne sont pas positionnés aléatoirement. Ces positions sont définies en fonction de la dynamique des cellules (division cellulaire, différenciation...), et en fonction de l'activité des gènes également. Certaines études montrent que la position du gène dans le noyau influence son état transcriptionnel (Tanabe *et al.* 2002; Mehta *et al.* 2010). En étudiant certaines espèces vertébrés, Tanabe et ses collaborateurs ont montré que les

chromosomes riches en gènes actifs sont localisés au centre nucléaire alors que ceux qui sont pauvres en gènes actifs, sont localisés vers les périphéries du noyau (Tanabe *et al.* 2002).

Par conséquent, la localisation des gènes dans le noyau est considérée comme un support épigénétique qui régule l'expression des gènes, cependant il reste le moins caractérisé à ce jour.

I.3.2. Etat chromatinien : Euchromatine / Hétérochromatine

Un état chromatinien d'un locus ou d'un gène n'est pas déterminé par un seul des quatre supports de l'information épigénétique, c'est le résultat des combinaisons de différentes marques épigénétiques au niveau des nucléosomes. La modification chimique de l'ADN associée à certaines modifications d'histones, ainsi qu'une variation du taux de certains ARN non codants (augmentation ou baisse), conduisent ensemble à former un état chromatinien, contrôlant l'accessibilité des facteurs de transcription du gène (Devaskar & Raychaudhuri 2007).

Nous distinguons deux types d'états chromatiniens les plus extrêmes : « hétérochromatine » pour désigner l'état de la chromatine fermé, compact et non permissif à la transcription et « euchromatine » quant à elle, désigne l'état chromatinien ouvert, relâché et permissif à la transcription (figure 9). Il existe différents niveaux d'hétérochromatisation et d'euchromatisation (Berger 2007; Mikkelsen *et al.* 2007; Roudier *et al.* 2011).

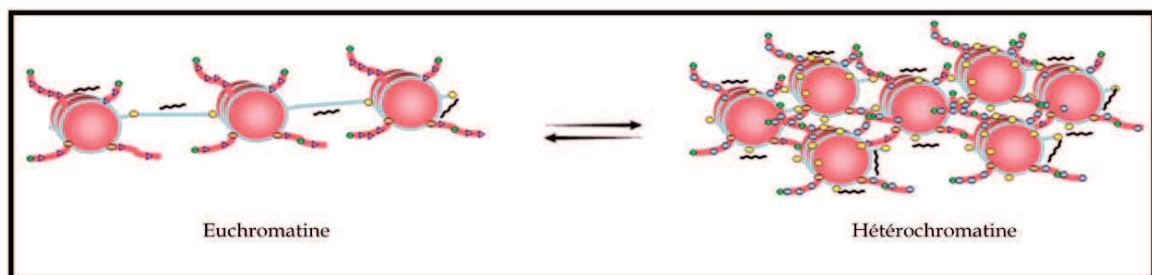


Figure 9 : La transition entre les deux états de la chromatine, euchromatine et hétérochromatine. Ce schéma représente la structure de l'euchromatine permissive à la transcription et l'hétérochromatine non permissive à la transcription.

Bien que les quatre bases porteuses de l'information épigénétique soient décrites, tous les mécanismes épigénétiques moléculaires ne sont pas encore étudiés à l'heure actuelle.

I.4. Modèle théorique : génétique / épigénétique / variabilité phénotypique

Dans des conditions environnementales particulières, des individus expriment de nouveaux phénotypes ($P=G \times E$) (Comstock & Moll 1963). La capacité des individus à produire différents phénotypes pourrait être expliquée par des mécanismes génétiques, ou autrement dit un phénotype différent correspond à une variation dans la séquence nucléotidique d'un gène. En revanche, des études et observations effectuées chez un bon nombre d'espèces telles que les mammifères et les plantes montrent que des mécanismes épigénétiques pourraient être aussi à la base de ces variations phénotypiques.

De leur côté, les mécanismes épigénétiques peuvent conférer aux génomes une plasticité phénotypique traduite par la capacité d'un seul génotype à exprimer un certain nombre de phénotypes en fonction des conditions environnementales (Godfrey *et al.* 2007). Les investigations moléculaires ont démontré que ces mécanismes épigénétiques sont basés globalement sur des modifications chimiques de la séquence de l'ADN et/ou des protéines histones qui induisent des changements dans la configuration des nucléosomes. La configuration des nucléosomes à son tour, affecte l'accessibilité de différents facteurs vers l'ADN et par conséquent, l'état chromatinien obtenu sera essentiel et déterminant de la transcription du gène. Donc, ces mécanismes épigénétiques seront à la base de l'expression de certains phénotypes sans altérer la séquence de l'ADN.

D'une manière générale, la mise en place d'un état chromatinien et son effet sur le phénotype semble être plus rapide, moins coûteuse et moins risqué qu'une mutation compte tenu de la réversibilité des marques épigénétiques.

Dans le but de mieux comprendre l'implication de l'épigénétique dans l'évolution adaptative, Pal et Miklos ont proposé un modèle théorique en 1999 (Pal & Miklos 1999). Ils suggèrent que dans les conditions normales, un caractère phénotypique X,

pourrait varier dans une population autour d'une valeur maximale tout en restant entre deux limites bien définies. En revanche, dans un environnement stressant, le spectre de la variabilité phénotypique augmente dans la population. Cela est traduit par un élargissement des intervalles des limites supérieures et inférieures (figure 10).

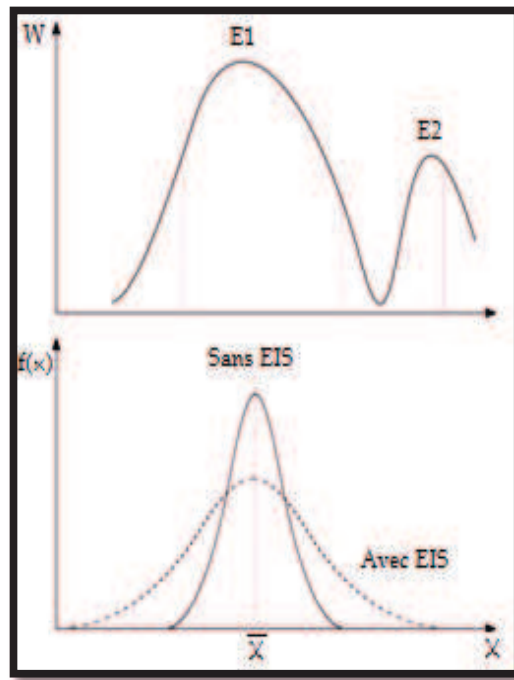


Figure 10 : Représentation schématique de la variabilité phénotypique selon les modèles théoriques proposés. La représentation de la fitness W (axe des ordonnées du haut), et de la fréquence du caractère phénotypique X (axe des ordonnées du bas) en fonction du caractère X dans une population (axe des abscisses). Les axes du haut : les deux différents environnements E1 et E2. Axes du bas : avec EIS, la variabilité phénotypique s'élargit suffisamment pour que des individus puissent survivre au nouvel environnement (EIS : Epigenetic Inheritance System). Adaptée à partir du modèle de Pal et Miklos (Pal & Miklos 1999).

Leur modèle est supporté très récemment par Klironomos et ses collègues (Klironomos *et al.* 2013). Selon les deux équipes (Pal & Miklos 1999; Klironomos *et al.* 2013) :

- a) Dans un premier temps, des variations épigénétiques pourraient être à l'origine d'une variabilité phénotypique. Cette variabilité pourrait être transmise à la descendance sans changement génétique.
- b) Au cours du temps, la sélection naturelle pourrait agir sur le phénotype ayant la meilleure valeur sélective.

c) L'étape finale sera la mise en place des changements génétiques.

Par conclusion, la génération de variants phénotypiques d'origine épigénétique, serait la première étape dans l'évolution adaptative, évoquant dès lors le phénomène d'assimilation génétique proposé par Waddington (Waddington 1953).

I.5. Interactions hôtes / parasites : bon modèle pour la biologie évolutive

Tous les êtres vivants sont concernés par des interactions hôte / parasite, en tant qu'hôte ou bien parasite (Thomas *et al.* 2010). L'interaction du parasite avec son hôte est obligatoire pendant l'intégralité ou bien une partie de son cycle de vie.

Les interactions hôte / parasite se caractérisent par un équilibre dynamique. En effet, au cours de certaines interactions, des molécules de surveillance de l'hôte pourraient reconnaître des protéines de surface du parasite. Cette reconnaissance permettrait l'hôte de se débarrasser du parasite. A son tour, le parasite va générer de nouvelles molécules afin d'éviter la défense de l'hôte jusqu'à ce que l'hôte, lui aussi génèrerait de nouvelles molécules de reconnaissance (Frank 2000).

Ces interactions constituent de bons modèles de coévolution où chacun des deux partenaires s'adapte continuellement aux conditions que lui impose l'autre. Dans de telles interactions, les pressions sélectives sont fortes conduisant à des évolutions qui seraient relativement rapides.

I.5.1. Modèle d'étude : *Biomphalaria glabrata* / *Schistosoma mansoni*

Dans notre laboratoire, nous disposons d'un modèle hôte / parasite. Le trématode parasite *Schistosoma mansoni* qui est responsable de la maladie de la bilharziose intestinale, avec son hôte intermédiaire le mollusque invertébré *Biomphalaria glabrata*.

I.5.1.1. La bilharziose : Deuxième endémie parasitaire mondiale

La bilharziose ou schistosomiase est une maladie parasitaire due à des vers plathelminthes, du genre *Schistosoma*. Cette maladie est classée comme deuxième endémie parasitaire mondiale après la malaria. Selon l'Organisation Mondiale de la Santé (OMS), environ 200 millions de personnes réparties dans 74 pays des régions

tropicales et subtropicales (figure 11), sont touchées par la bilharziose (Chitsulo *et al.* 2000; Gryseels *et al.* 2006). Les trois espèces les plus pathogènes de Schistosomes sont *S. mansoni* (localisée principalement en Afrique et Amérique du sud) et *S. japonicum* (présente au sud et à l'est de l'Asie) provoquant les bilharzioses hépato-splénique et intestinale, et *S. haematobium* (présente principalement en Afrique) et responsable de la bilharziose urinaire (Gryseels *et al.* 2006). C'est une maladie qui provoque 200 000 à 300 000 morts par an selon l'OMS (Gryseels *et al.* 2006), et caractérisée par des interactions évolutives et très dynamiques entre les schistosomes et leurs hôtes. Il serait donc important d'étudier les bases moléculaires à l'origine de l'évolution adaptative des schistosomes.

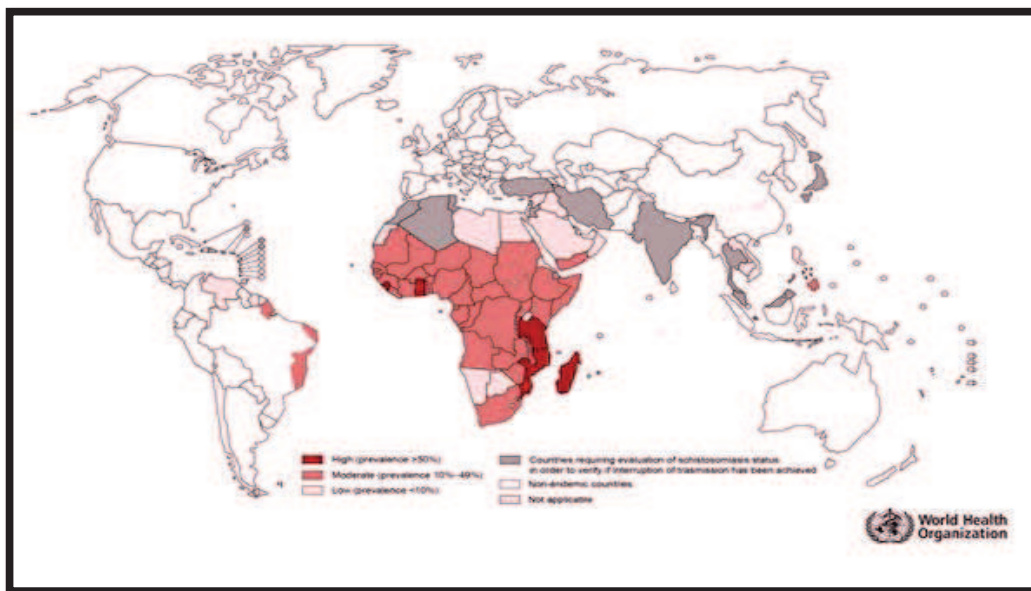


Figure 11: La répartition mondiale de la bilharziose. Source OMS, http://gamapserver.who.int/mapLibrary/Files/Maps/Schistosomiasis_2012.png.

I.5.1.2. Cycle de vie de *S. mansoni*

S. mansoni se caractérise par un cycle de vie qui est complexe, nécessitant le passage par deux hôtes obligatoires et successifs. L'hôte intermédiaire qui est un mollusque planorbe d'eau douce du genre *Biomphalaria*, dans lequel s'effectue la multiplication asexuée. L'hôte définitif vertébré, dans lequel s'effectue la reproduction sexuée de *S. mansoni*, est primate ou rongeur (figure 12).

Les schistosomes atteignent leur stade adulte dans l'hôte vertébré. Ils vivent et s'accouplent dans le système veineux. La femelle, plus étroite que le mâle, pond les œufs dans les veinules mésentériques. Les femelles peuvent pondre 200 à 300 œufs par jour. A l'aide de leur éperon, une partie des œufs perforent la paroi intestinale et sont éliminées par les selles. Le reste des œufs sont transportés par le flux sanguin et vers le foie, où ils s'accumulent et provoquent des inflammations responsables de la pathologie (fibrose hépatique). Au contact de l'eau douce, les œufs éclosent et libèrent une larve ciliée, le miracidium (premier stade larvaire des schistosomes).

Le miracidium nage activement pour trouver son hôte intermédiaire, le mollusque d'eau douce du genre *Biomphalaria*. Dans le mollusque, le schistosome se multiplie d'une manière asexuée. Les miracidia se transforment donc en sporocystes primaires (Sp1) ou sporocystes mères, qui à leur tour se développent asexuellement en sporocystes secondaires filles (Sp2). Ces dernières se libèrent pour s'installer dans l'hépatopancréas du mollusque. Dans les sporocystes se développe une nouvelle forme larvaire, les cercaires (deuxième stade larvaire des schistosomes) (Rollinson & Simpson 1987).

Les cercaires, munies d'une queue, traversent le pied du mollusque pour atteindre l'eau. La pénétration d'un seul miracidium dans le mollusque peut être à l'origine de milliers de cercaires (Theron 1986). Il est important de noter que l'infestation du mollusque par un seul miracidium (infestation monomiracidiale), conduit à la production d'une population clonale de cercaires par multiplication asexuée, ayant un sexe unique et un seul génotype. La durée du cycle dans le mollusque depuis la pénétration du miracidium, jusqu'à la libération des cercaires est de 4 à 5 semaines.

La cercaire, larve nageuse très active, est responsable de l'infestation de l'hôte définitif (primate ou rongeur) au moment où ils se baignent dans l'eau douce. Seul le corps de la cercaire pénètre alors que la queue se détache et reste dans le milieu extérieur. Dans le corps humain (ou des rongeurs), la cercaire suit la circulation sanguine pour atteindre le cœur vingt-quatre heures après la pénétration (Standen 1953). De là, la cercaire se développe en schistosomule (avant l'arrivée au stade adulte) et continue son chemin vers les poumons et puis vers le foie.

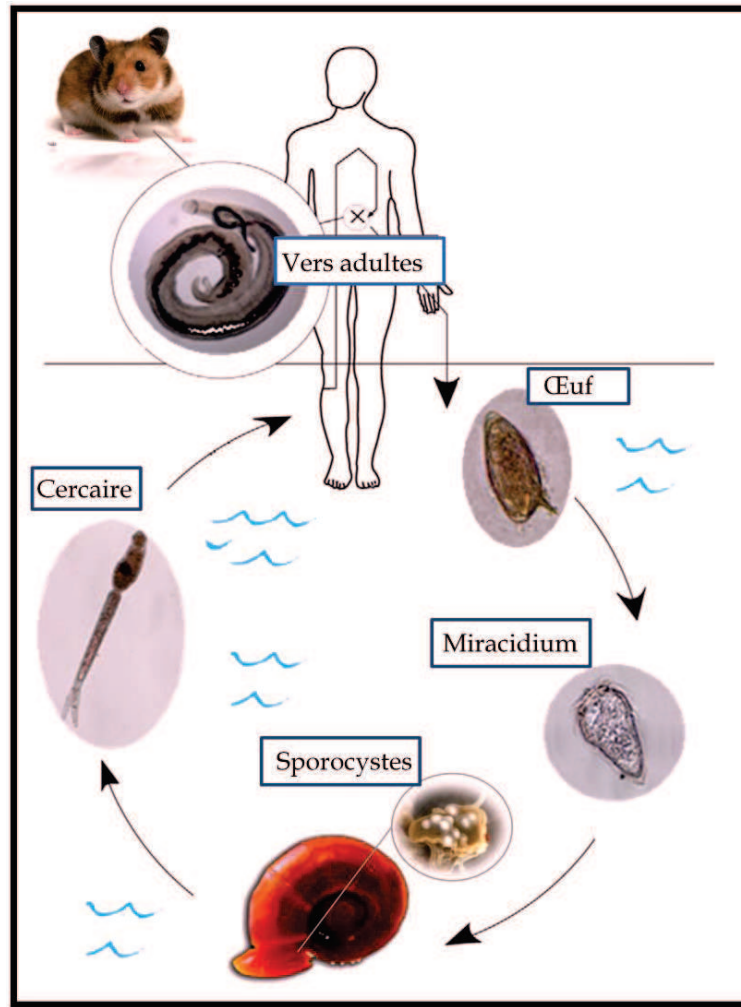


Figure 12: Cycle de vie de *S. mansoni*. Après la reproduction sexuée des schistosomes dans l'hôte définitif (primate ou rongeur), les œufs du parasite sont émis dans l'eau. Une fois au contact de l'eau, les œufs éclosent pour libérer les miracidia. Ces larves nagent activement pour rencontrer leur hôte intermédiaire, mollusque du genre *Biomphalaria*, et y pénétrer. Dans le mollusque, les miracidia se développent par multiplication asexuée en sporocystes, qui à leur tour vont générer des larves libres ou cercaires. Les cercaires pénètrent dans le corps humain (ou des rongeurs) par la peau, où ils atteignent le stade adulte, se reproduisent et émettent des œufs.

Arrivés au foie, les schistosomules gagnent la veine porte et deviennent sexuellement matures (Rollinson & Simpson 1987). La durée du cycle depuis la pénétration des cercaires dans l'hôte définitif jusqu'à l'apparition des premiers œufs dans les selles est au minimum de 40 à 50 jours.

Jusqu'à présent, la bilharziose était considérée comme une maladie tropicale et subtropicale. Cependant, différents cas de bilharziose signalés en Europe au mois d'avril de cette année 2014, nous force à reconsidérer cette vision. Les recherches menées à l'heure actuelle, dans lesquelles notre laboratoire est impliquées, ont mis en évidence une émergence de la bilharziose en Corse et possiblement en Europe continentale (Berry *et al.* 2014). En effet, les mollusques d'eau douce, hôtes intermédiaires des schistosomes et plus particulièrement, la souche *S. haematobium*, sont déjà présents en Europe du sud (Portugal, Espagne et Corse) (Doby *et al.* 1965; Perez-Quintero *et al.* 2004). Certainement, le réchauffement climatique ou même les facteurs anthropiques tels que les mobilités des populations, sont à l'origine de la dispersion de toutes sortes de maladie partout sur le globe terrestre.

La forte capacité d'adaptation des schistosomes, suite à des fluctuations environnementales naturelles ou humaines, est déjà connue (Zhou *et al.* 2008; Polley *et al.* 2010). En effet, les études phylogénétiques ont montré que les Schistosomes seraient originaires d'Asie, et auraient colonisé l'Afrique il y a une quinzaine de millions d'années. Ces études ont mis en évidence un transfert récent de l'espèce *S. mansoni*, notre espèce modèle au laboratoire, de l'Afrique vers l'Amérique du sud et les Caraïbes. Ce transfert se serait effectué suite à la traite des esclaves vers le XVI^{ème} et XVII^{ème} siècle (Morgan *et al.* 2001; Lockyer *et al.* 2003). L'arrivée des schistosomes sur un nouveau continent, nécessitait une adaptation à de nouveaux hôtes intermédiaires. *S. mansoni* a montré une capacité d'adaptation élevée aux espèces de mollusque d'eau douce présente sur le continent du nouveau monde du genre *Biomphalaria* (Morgan *et al.* 2001). Il a en effet été montré que les espèces *B. straminea* (Belén Par, Brésil), *B. tenagophila* (Asuncion, Brésil), *B. amazonica* (Amazonas, Brésil) et d'autres espèces de l'Amérique latine, pouvaient être des hôtes intermédiaires de *S. mansoni* sur ce continent (DeJong *et al.* 2001).

Les schistosomes s'adaptent non seulement aux nouveaux environnements et aux nouveaux hôtes intermédiaires, mais ils ont montré également des cas de résistance aux différentes molécules chimiques anti-helminthes développées jusqu'à présent (Julie 2012).

A l'heure actuelle, le traitement de choix contre la bilharziose est le praziquantel (PZQ). Découvert en 1977, le PZQ a montré une efficacité non négligeable par rapport aux autres traitements développés (Thomas & Gönner 1977). En effet, il endommage le tégument des vers adultes de toutes les espèces de schistosomes (Gryseels *et al.* 2006). Cependant, il agit peu ou pas sur les stades œuf ou larvaires du schistosome. L'avantage de l'utilisation du PZQ c'est qu'il est administré en une seule dose orale et qu'il provoque peu ou pas d'effets secondaires chez les patients traités (Fenwick *et al.* 2003).

Or, dans les zones endémiques telles que l'Égypte, les gens sont souvent confrontés à des infestations successives par des schistosomes, et l'administration de plusieurs doses de PZQ serait essentielle. Cependant, l'utilisation intensive et répétée du PZQ mène au développement des cas de résistance (Ismail *et al.* 1999; Gryseels *et al.* 2001; Doenhoff *et al.* 2008).

L'avantage de notre laboratoire est d'avoir accès expérimentalement aux différents stades du développement du parasite et de ses hôtes. Nous disposons de différentes souches du parasite *S. mansoni*, ainsi que de différentes souches du mollusque *B. glabrata*, correspondant à des isolats géographiques différents. Les cycles de vie de ces différentes interactions sont maintenus, aussi bien sur l'hôte intermédiaire que sur l'hôte définitif.

1.5.2. La théorie de la Reine Rouge

L'environnement est en constante variation et l'adaptation des organismes à l'environnement est continue. De ce fait, Van Valen a proposé la théorie de la Reine Rouge en 1974 (Van Valen 1974). Il s'est inspiré d'une image du roman de Lewis Carroll, *De l'autre côté du miroir et ce qu'Alice y trouva*, dans lequel Alice demanda à la Reine Rouge : « *Nous courons vite et le paysage ne change pas?* » et la reine répondit : « *Nous courons pour rester à la même place* » (figure 13). Ceci illustre parfaitement les espèces qui évoluent pour s'adapter à un environnement changeant. Cette théorie s'applique aux interactions hôtes / parasites. En effet, ces interactions se caractérisent par des sélections successives des hôtes qui développent des mécanismes de défense

contre les parasites d'un côté, et des parasites qui échappent aux moyens de défense de l'hôte d'un autre côté. Cette coévolution hôte / parasite explique une course aux armements, qui nécessite la production de variants phénotypiques aussi bien de la part de l'hôte que de la part du parasite pour garder l'équilibre progressif de l'interaction (Frank 2000).

Les interactions hôte/ parasite constituent de bons modèles pour l'étude de la biologie évolutive. Il est important de noter que l'hôte constitue l'environnement du parasite au cours de son développement, et à son tour le parasite constitue l'environnement stressant de l'hôte. Par conséquent, les deux partenaires produisent de nouveaux phénotypes dans le but de se contrecarrer l'un l'autre. Les mécanismes à la base de la production des phénotypes adaptatifs peuvent être génétiques et/ou épigénétiques.

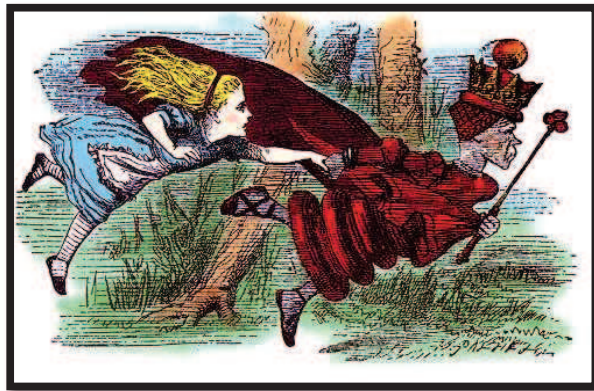


Figure 13 : Alice et la Reine Rouge. Image tirée du roman « *De l'autre côté du miroir et ce qu'Alice y trouva* » de Lewis Carroll (1871). Le paysage désigne l'environnement et la course désigne l'évolution.

A l'heure actuelle, les mécanismes à la base de la défense de l'hôte et de l'infectivité du parasite ne sont pas clairs.

Deux différentes écoles de pensées ont émergé pour expliquer les mécanismes à la base de la coévolution rapide entre hôtes / pathogènes. L'une des écoles a proposé l'hypothèse « *Gene for gene* », et la deuxième est plutôt pour l'hypothèse du « *Matching phenotype* ».

I.5.2.1. « *Gene for gene* »

L'hypothèse « *gene for gene* » a été proposée pour la première fois par le botaniste Flor (1900-1991) il y a 50 ans environ, après ses observations sur des interactions plante / pathogène (Flor 1971). C'est un concept basé sur la variabilité génétique des hôtes et des pathogènes. Pour chaque gène de résistance (*R*) chez l'hôte correspond un gène d'avirulence (*Avr*) chez le parasite (Van Der Biezen & Jones 1998).

Cette théorie peut être illustrée par un modèle de récepteur – ligand : lors d'une infection d'un hôte par un parasite à gène de virulence (*V*), cet hôte va combattre le parasite par des produits du gène de résistance (*R*). Une fois que l'hôte reconnaît les produits du gène de virulence (*V*), il doit impérativement se défendre contre son pathogène, et met en place un mécanisme de défense, le gène (*V*) devient alors avirulent (*Avr*). Pour l'intérêt et le développement du parasite, celui-ci va contourner la machinerie de l'hôte en modifiant ses gènes (*Avr*) par des gènes (*V*), et ainsi de suite. Ce sont les bases génétiques (un ou plusieurs gènes) qui déterminent si l'interaction entre l'hôte et le pathogène est compatible ou incompatible (Prell & Day 2001).

Ce modèle a été décrit dans plusieurs cas d'interactions plantes / pathogènes (Van Der Biezen & Jones 1998), cependant les mécanismes à la base de la variation de ces gènes (mutations, éléments transposables...) ne sont pas tous clairs.

I.5.2.2. « *Matching phenotype* »

Basch était le premier à proposer le modèle du « *matching phenotype* » en 1975, dans une interaction mollusque / trématode (Basch 1975). Contrairement à l'hypothèse « *Gene for gene* », cette hypothèse implique une concordance des phénotypes spécifique entre les différents facteurs de résistance/susceptibilité chez l'hôte, et les différents facteurs de virulence/infectivité chez le parasite. Il est important de noter que les interactions mollusque / trématode sont des interactions spécifiques (Adema *et al.* 1997a).

Les principales différences entre les deux hypothèses :

(i) L'hypothèse « *Gene for gene* » implique un seul gène de l'hôte et un seul gène du parasite. En revanche, le modèle du « *matching phenotype* » repose sur une interaction qui a lieu entre le phénotype de l'hôte avec celui du parasite.

(ii) La validation de l'hypothèse « *Gene for gene* » serait plus difficile dans les modèles où les organismes possèdent le même génotype que celle du « *matching phenotype* ». En effet, les phénotypes tiennent en compte tous les mécanismes générateurs de la variabilité. L'origine du phénotype pourrait être génétique et / ou épigénétique.

Cette théorie a trouvé des validations expérimentales suite aux travaux effectués par Théron et Coustau en 2005 sur le modèle d'interaction entre *Schistosoma* et *Biomphalaria*. Ce travail a montré que la compatibilité s'opérait à l'échelle des individus, d'une manière spécifique à chaque interaction entre un parasite et un hôte. Ils ont démontré que le succès ou l'échec d'une infestation ne dépend pas d'un mécanisme de résistance/virulence spécifique entre le mollusque et le parasite, mais ceci est le résultat d'une concordance ou pas entre les phénotypes du parasite et de l'hôte (Théron & Coustau 2005). Une représentation du modèle « *matching phenotype* » est illustrée dans la figure 14.

Ainsi, les résultats expérimentaux obtenus jusqu'à présent, sont en faveur de l'hypothèse du « *matching phenotype* » ou bien la concordance des phénotypes, sur notre modèle. En effet, cette hypothèse implique une concordance spécifique entre les différents facteurs de résistance/susceptibilité chez l'hôte, et les différents facteurs de virulence/infectivité chez le parasite.

En ce sens, la théorie de la Reine Rouge s'applique parfaitement à notre modèle qui nécessite un équilibre entre les mécanismes de défense de l'hôte et le pouvoir infestant du parasite. Par ailleurs, cette théorie prédit que l'équilibre des mécanismes d'attaque et de défense des deux partenaires favorise la diversification (Frank 2000). Nous verrons plus loin que cette diversification trouve aussi des validations sur notre modèle (au niveau moléculaire).

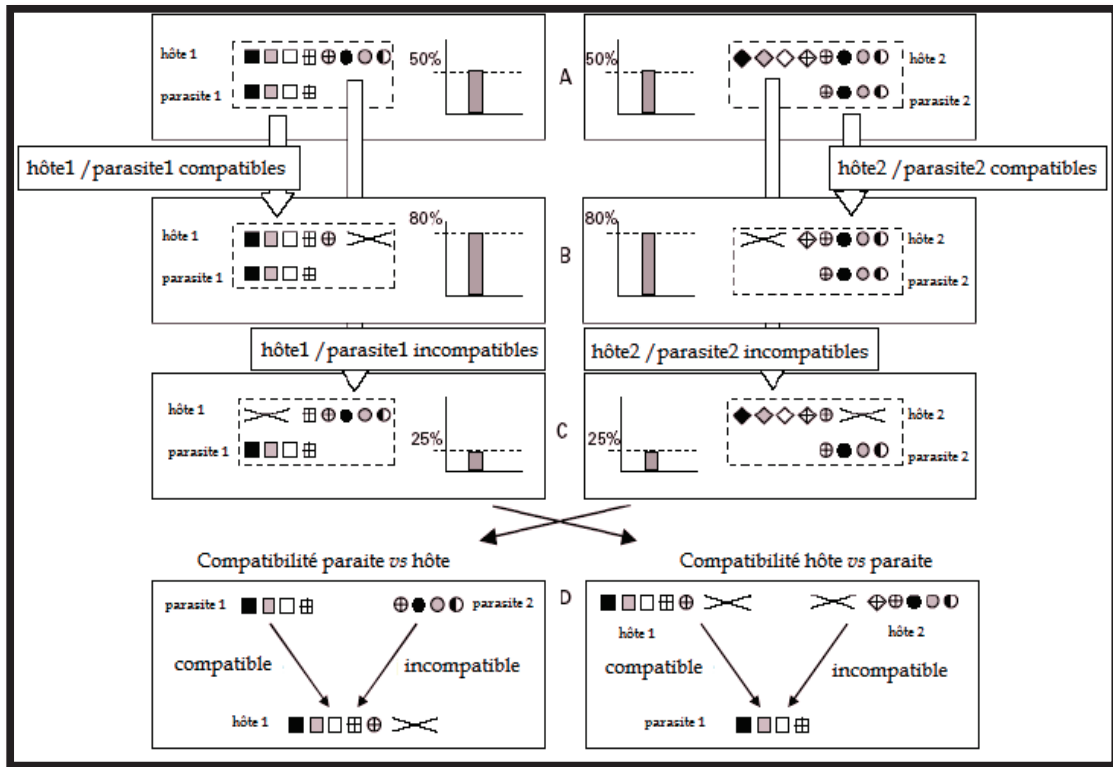


Figure 14 : Modèle du « matching phenotype ». Représentation schématique de la sélection des couples hôte / parasite compatibles et incompatibles. Deux souches différentes d'une même espèce d'hôte et deux souches d'une même espèce de parasite sont schématisées selon l'hypothèse du « matching phenotype ». Les symboles représentent la diversité phénotypique de compatibilité entre les hôtes et les parasites. (A) 50 % des phénotypes des souches d'hôtes concordent avec les phénotypes des souches du parasite. (B) Sélection des couples hôte / parasite compatibles où la concordance des phénotypes est élevée. (C) Sélection des couples hôte / parasite incompatibles où la concordance est faible. (D) La compatibilité est réciproque : deux différentes souches du parasite, une est compatible et l'autre est incompatible vis-à-vis de la même souche de l'hôte. Et en parallèle, deux différentes souches de l'hôte, l'une est compatible et l'autre est incompatible vis-à-vis de la même souche du parasite. Adaptée à partir de (Théron & Coustau 2005).

1.5.3. Polymorphisme de compatibilité

L'hypothèse du « matching phénotype » appliquée à l'interaction entre le parasite *S. mansoni* avec son hôte intermédiaire *B. glabrata* se traduit par un polymorphisme de compatibilité en population naturelle. Le polymorphisme de compatibilité implique la présence des phénotypes hôte / parasite compatibles et d'autres incompatibles. Ceci a été mis en évidence par André Théron et ses collaborateurs (Théron *et al.* 2014).

Par ailleurs, il est possible de reproduire ce phénomène au laboratoire en utilisant deux souches différentes du parasite *S. mansoni*, qui sont soit compatible ou incompatible vis-à-vis d'une même souche de mollusque *B. glabrata* (Théron & Coustau 2005). Dans l'interaction compatible, le parasite se développe normalement dans l'hôte. Alors que dans l'interaction incompatible, le parasite est soit directement reconnu par l'hôte, encapsulé et éliminé, soit il se développe mais avec de très faibles prévalences (voir définition plus tard).

Des approches moléculaires globales ont été menées dans notre laboratoire, et ont mis en évidence des éléments moléculaires clés dans l'interaction mollusque / parasite. Différents candidats ont été identifiés, parmi eux des molécules hautement diversifiées aussi bien de la part de l'hôte que de la part du parasite. Les études des bases moléculaires du polymorphisme de compatibilité ont fait l'objet de deux thèses :

- 1) Bases génétiques du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata* par Emmanuel Roger. Au cours de sa thèse, il a identifié et caractérisé des *Schistosoma mansoni* Polymorphic Mucins (*SmPoMucs*), molécules hautement diversifiées de *S. mansoni* (Emmanuel 2008). Nous reviendrons à la présentation de ces *SmPoMucs* plus loin dans l'introduction.
- 2) Bases moléculaires du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata* par Yves Moné. Il a mis en évidence les partenaires des *SmPoMucs* par des expériences d'immunoprécipitation. Il a notamment montré une interaction des *SmPoMucs* avec d'autres molécules hautement diversifiées produites par le mollusque, les Fibrinogen-Related Proteins (FREPs) (Yves 2011).

Au cours de ces travaux, E. Roger et Y. Moné ont montré que la diversification moléculaire est favorisée dans le cas d'une interaction hôte invertébré / trématode. Par ailleurs, ils ont confirmé les prédictions de la théorie de la Reine Rouge, en mettant en évidence des molécules diversifiées qui interagissent entre elles dans un contexte de course aux armements entre l'hôte et le parasite.

Ma thèse s'inscrit dans cette thématique d'interaction et du polymorphisme de compatibilité entre *B. glabrata* et *S. mansoni*. En effet, je me suis attachée à déterminer si l'épigénétique pouvait être une des composantes à l'origine de la variabilité phénotypique dans cette interaction.

Mes travaux sont divisés en deux volets : (i) Du côté du parasite *S. mansoni*, je me suis intéressée aux mécanismes moléculaires à la base du processus de diversification des molécules *SmPoMucs*. Les *SmPoMucs* étaient les marqueurs moléculaires de choix pour cette étude pour deux principales raisons : elles sont hautement diversifiées et joueraient un rôle clé dans le polymorphisme de compatibilité. (ii) Du côté de l'hôte *B. glabrata*, je me suis investie dans la caractérisation des bases épigénétiques.

I.5.3.1. Polymorphisme de compatibilité – traits de vie

André Théron et ses collaborateurs ont mis en évidence les degrés de susceptibilité des différentes souches de *B. glabrata* vis-à-vis de différentes souches de *S. mansoni* (Theron *et al.* 2014). En parallèle, ils ont caractérisé l'infectivité de différentes souches de *S. mansoni* vis-à-vis de différentes souches de *B. glabrata*. Les traits de vie mesurés sont les prévalences et les intensités.

La prévalence désigne la proportion de mollusques infestés par rapport au nombre total de mollusques qui ont été exposés au parasite. L'intensité correspond au nombre de sporocystes primaires développés à l'intérieur du mollusque infesté. Dans le but de tester les prévalences et les intensités, chaque mollusque est exposé à 20 miracidia d'une même souche de *S. mansoni*.

Nous avons retenu le nombre 20 suite aux résultats obtenus au laboratoire par la même équipe. En effet, afin d'étudier les variations de prévalences, ils ont exposé chaque individu mollusque à différentes doses de 1, 10, 20, 30 et 50 miracidia. Les résultats ont montré qu'à partir de 20 miracidia le taux d'infection n'augmente plus (Theron *et al.* 2014). La dose de 20 miracidia contient donc toute la diversité phénotypique potentiellement susceptible de présenter au moins un phénotype qui « *matche* » avec le phénotype du mollusque exposé.

La figure 15 montre un polymorphisme de compatibilité de deux différents couples mollusque / parasite provenant de deux isolats géographiques différents. Les souches brésiliennes *B. glabrata* BRE / *S. mansoni* BRE et les souches guadeloupéennes *B. glabrata* GUA / *S. mansoni* GH2.

Le cycle de vie de ces deux souches est maintenu au laboratoire depuis 1970 pour les souches brésiliennes et 1983 pour les souches guadeloupéennes (communication personnelle avec André Théron). La souche brésilienne du parasite *Sm* BRE est 100% compatible avec son hôte sympatrique *Bg* BRE, c.à.d. tous les mollusques exposés aux parasites sont infestés avec une intensité moyenne de 7.12 ± 0.25 . Alors qu'elle est moins compatible avec son mollusque allopatrique *Bg* GUA, où 80% de *Bg* GUA sont infestés par *Sm* BRE avec une intensité moyenne de 3.03 ± 0.27 . La souche guadeloupéenne *Sm* GH2 présente différents seuils d'incompatibilité avec les différentes souches de mollusques. *Sm* GH2 infeste son mollusque sympatrique *Bg* GUA avec une prévalence de 40% et une intensité moyenne de 2.6 ± 0.34 et son hôte allopatrique *Bg* BRE avec une prévalence 5% et une intensité moyenne de 1 ± 0.00 . Ces valeurs sont montrées dans les travaux de (Theron *et al.* 2014), et au cours de mes travaux de thèse.

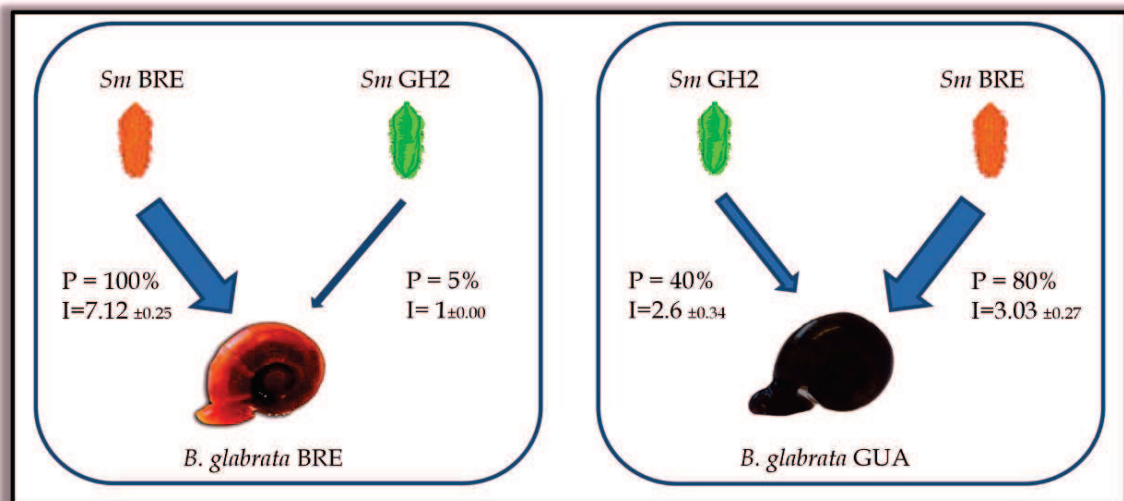


Figure 15 : Polymorphisme de compatibilité du couple *B. glabrata* /*S. mansoni* de deux souches d'isolats géographiques différents. Le pourcentage de prévalences (P) et les intensités (I) sont indiqués pour chacune des combinaisons sympatriques *Bg* BRE /*Sm* BRE et *Bg* GUA /*Sm* GH2, et allopatriques *Bg* BRE /*Sm* GH2 et *Bg*

GUA/*Sm* BRE. Chaque infestation a été effectuée en exposant 20 miracidia face à un individu mollusque.

Ce polymorphisme de compatibilité des traits de vie vis-à-vis du mollusque repose surtout sur des bases moléculaires de reconnaissance de la part de l'hôte et de la part du parasite.

1.5.3.2. Polymorphisme de compatibilité moléculaire

Le polymorphisme de compatibilité caractéristique de l'interaction *B. glabrata* / *S. mansoni* est certainement supporté par des bases moléculaires clés qui maintiennent la dynamique de l'interaction. Selon l'hypothèse du « *matching phenotype* », les hôtes et les parasites seront munis de molécules de reconnaissance, qui devraient être diversifiées permettant la concordance ou pas de l'interaction. Des études protéomiques comparatives effectuées au laboratoire entre les deux différentes souches du parasite, brésilienne *Sm* BRE et guadeloupéenne *Sm* GH2, ont mis en évidence des candidats moléculaires de la compatibilité chez le parasite.

Parmi les candidats, la famille multigénique *SmPoMucs* code des protéines très polymorphes, fortement glycosylées de type mucine (Roger *et al.* 2008c). Ce sont des variants moléculaires polymorphes sécrétées surtout aux stades miracidium, stade infestant le mollusque (figure 16 A). Ces *SmPoMucs* sont sécrétées au niveau des glandes apicales du miracidium (figure 16 B), glandes sécrétrices des substances protéolytiques favorisant la pénétration du parasite au travers du tégument de l'hôte (Roger *et al.* 2008a). *SmPoMuc* est une famille multigénique qui code pour au moins 10 gènes divisés en 4 groupes (groupe 1, groupe 2, groupe 3 et groupe 4). Chaque groupe, à son tour est divisé en sous-groupes. Ces gènes codent pour des mucines, fortement glycosylées dans leur région N-terminale composée d'un domaine pour Variable Number of Tandem Repeats (VNTR). Ce domaine est composé de 9 acides aminés organisés en tandem ou unités répétées, composées essentiellement par les trois acides aminés : sérine (Ser), thréonine (Thr) et proline (Pro).

Nous distinguons trois différents types d'unités répétées r1, r1' et r2 qui peuvent être répétées de 1 à 55 fois (figure 17 A). Cette région N-terminale constituée

essentiellement des unités répétées, se caractérise non seulement par un polymorphisme de séquence, mais aussi par un polymorphisme de O-glycosylation en fonction du nombre des unités répétées et de la position des acides aminés (Roger *et al.* 2008b). Ce qui augmente aussi le degré du polymorphisme. La glycosylation des *SmPoMucs* chez le parasite jouerait le rôle du déclenchement des mécanismes immunitaires chez l'hôte et la protection du parasite (Theodoropoulos *et al.* 2001). Le polymorphisme des *SmPoMucs* s'exprime aussi bien au niveau individuel de chaque des souches du parasite (figure 16 C), qu'au niveau populationnel (figure 16 D).

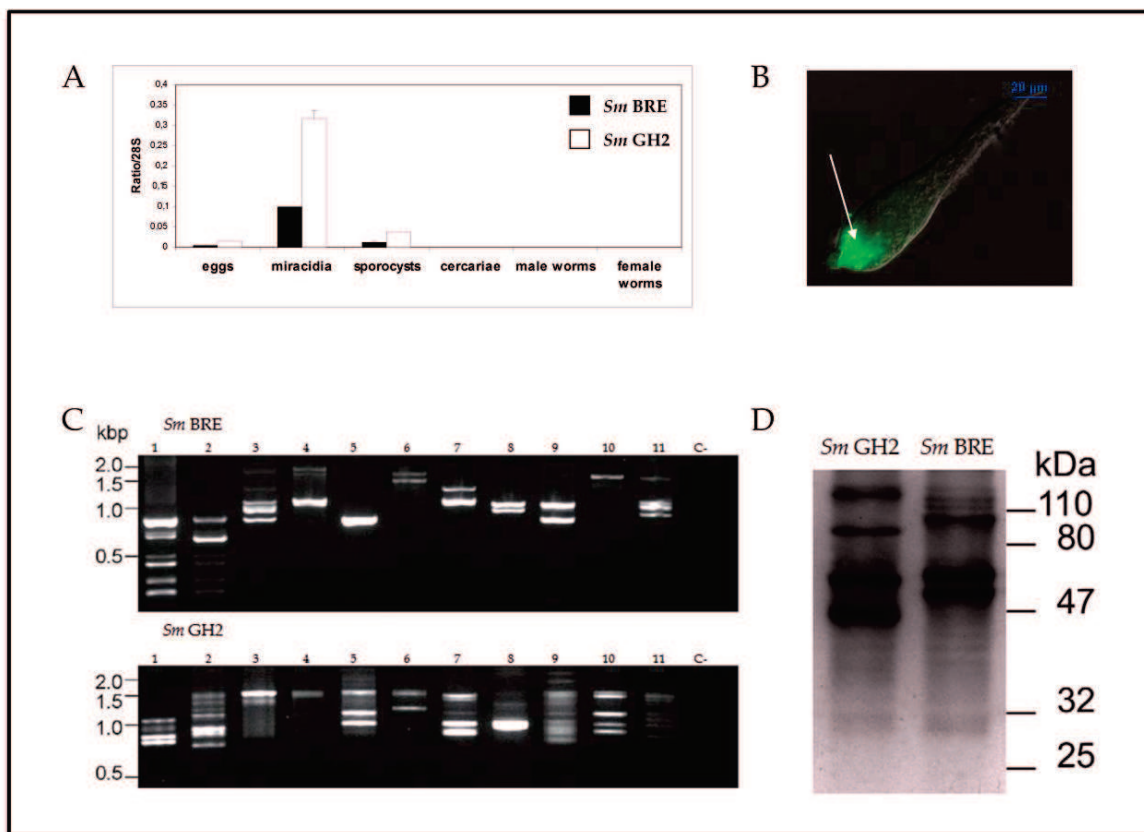


Figure 16 : Localisation et Polymorphisme d'expression des *SmPoMucs* chez les deux différentes souches de *S. mansoni*. (A) Cette figure représente le ratio de taux des transcrits de *SmPoMucs* par rapport à celui de l'ARNr 28S. *Sm BRE* et *Sm GH2* expriment les *SmPoMucs* essentiellement au stade miracidia. (B) Immunodétection des *SmPoMucs* dans la glande apicale du *Sm GH2*, utilisant un anti-*SmPoMucs* IgG (Roger *et al.* 2008a). (C) Profil de transcription individuel de *SmPoMucs* par RT-PCR, chez 11 individus *Sm BRE* et 11 individus *Sm GH2* (Roger *et al.* 2008b). (D) Profil d'expression protéique de *SmPoMucs* par Western blot, chez *Sm BRE* et *Sm GH2* utilisant anti-*SmPoMucs* (Moné *et al.* 2010).

Cependant la région C-terminale est relativement conservée entre les différents gènes codant les *SmPoMucs*, et par conséquent le polymorphisme des *SmPoMucs* repose surtout sur la région N-terminale.

Un des éléments marquant du polymorphisme des *SmPoMucs* est le gène *SmPoMuc 3.1 (r1r2)*, où un effet de « combinatoire » a été observé. En effet, ce polymorphisme de combinatoire se traduit par la présence des deux types d'unités répétées r1 et r2 dans le même domaine VNTR d'un gène appartenant au groupe 3 (Roger *et al.* 2008b). Ce variant *SmPoMuc 3.1 (r1r2)*, montre une expression différentielle entre les souches de parasites compatibles et incompatibles.

En effet, ce gène est transcrit chez la souche *Sm GH2* alors qu'il est peu ou pas transcrit chez la souche *Sm BRE*. Ceci est illustré dans la figure 17 A où le cDNA du *SmPoMuc 3.1 (r1r2)* est amplifié surtout chez la souche *Sm GH2* (Roger *et al.* 2008b). Des résultats d'expression effectués au laboratoire (non publiés), montrent que ce gène est beaucoup plus exprimé chez les souches du parasite incompatibles par rapport aux souches compatibles. Ici, les souches sont appelés compatibles ou incompatibles vis-à-vis de leur compatibilité avec le mollusque souche BRE (*Bg BRE*) (figure 17 B).

Les *SmPoMucs* constituent de très bons candidats moléculaires pour le polymorphisme de compatibilité, et potentiellement des molécules de reconnaissance par le système immunitaire de l'hôte invertébré *B. glabrata*. Le polymorphisme de transcription de *SmPoMuc 3.1 (r1r2)*, a fait de ce gène le candidat de choix pour nos études.

Des études menées chez le mollusque *B. glabrata*, ont mis en évidence une famille multigénique, codant pour des immunoglobulines hautement diversifiées de type lectine, les FREPS pour Fibrinogen-Related Proteins (Adema *et al.* 1997b). Les FREPs constituent une famille multigénique composée d'au moins pour 14 gènes (FREP 1 à FREP 14). Chacune des FREPs possède un ou deux domaines de type immunoglobuline en N-terminal et un domaine fibrinogène en C-terminal. Des études fonctionnelles basées sur l'ARN interférent montre leur implication dans la

reconnaissance et l'élimination du parasite (Hanington *et al.* 2012), jouant le rôle de défense contre les pathogènes.

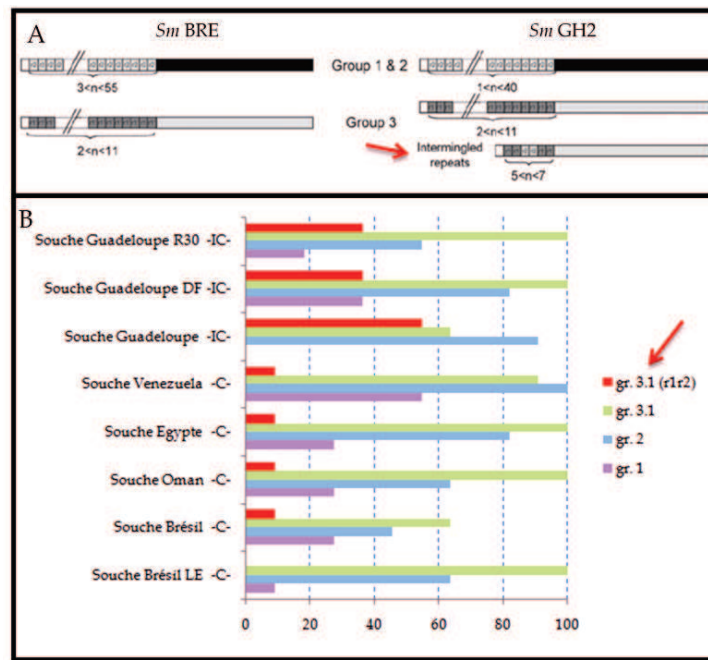


Figure 17: Polymorphisme transcriptionnel de la combinatoire *SmPoMuc* 3.1 (r1r2). (A) Les trois types d'unités répétées identifiés dans les cDNAs des *SmPoMuc*s (r1, r1' et r2) (Roger *et al.* 2008b). Les cDNA de *SmPoMuc* 3.1 (r1r2) sont détectés uniquement chez la souche *Sm* GH2 (flèche rouge). (B) Résultat de transcription des *SmPoMuc*s groupes 1, 2, 3.1 et 3.1 (r1r2), effectué sur 11 individus de chaque souche de *S. mansoni*, IC : incompatible et C : compatibles vis-à-vis du *Bg* BRE. En abscisse, c'est le pourcentage de la représentativité de chacun des groupes de *SmPoMuc*s dans la population de 11 individus. *SmPoMuc* 3.1 (r1r2) est peu ou pas du tout exprimé chez les souches compatibles (données du laboratoire non publiées).

A l'instar des *SmPoMuc*s, les FREPs sont des molécules très diversifiées au niveau individuel (Zhang *et al.* 2004; Hanington & Zhang 2010). Ces antigènes diversifiés interagissent avec les *SmPoMuc*s du parasite formant des complexes potentiellement impliqués dans la compatibilité / incompatibilité de l'interaction hôte / parasite (Moné *et al.* 2010). Cette reconnaissance des deux répertoires des molécules hautement diversifiées de la part de l'hôte et de la part du parasite confirme l'hypothèse du « *matching phenotype* », élément clé pour le succès / échec de l'infestation, pour revue voir (Mitta *et al.* 2012).

I.5.4. Mécanismes de diversification des éléments clés de l'interaction

Dans ce processus de course aux armements, les deux partenaires l'hôte *B. glabrata* et le parasite *S. mansoni* disposent d'une plasticité phénotypique moléculaire essentielle pour l'interaction. Le parasite exprime un répertoire de protéines polymorphes, les *SmPoMucs*, qui interagissent à leur tour avec un répertoire de molécules diversifiées chez le mollusque, les FREPs. Cette interaction FREPs / *SmPoMucs*, avec l'intervention d'autres molécules pourrait participer au phénomène de compatibilité / incompatibilité de l'interaction hôte / parasite (Mitta *et al.* 2012).

La particularité de ces molécules hautement diversifiées est qu'elles sont codées par un faible nombre de gènes. Le nombre de gènes ne pourrait pas expliquer toute la variabilité de transcrits obtenus aussi bien pour les *SmPoMucs* que pour les FREPs. La compréhension de ces phénomènes de diversification constitue un enjeu scientifique majeur d'un point de vue évolutif. En effet, la découverte d'autres molécules hautement diversifiées chez les invertébrés (Adema *et al.* 1997b) et leur implication dans la réponse immunitaire innée ont secoué le paradigme généralement admis du système immunitaire exclusivement non spécifique chez les invertébrés (Zhang *et al.* 2004). Comprendre les origines de cette diversification moléculaire est essentielle pour bien comprendre les phénomènes de coévolution.

Chez *B. glabrata*, les études effectuées jusqu'à l'heure actuelle montrent que la diversification de la FREP 3 chez *B. glabrata* est à l'origine des modifications somatiques aléatoires à travers des mutations et des conversions géniques (Zhang *et al.* 2004).

En ce qui concerne les *SmPoMucs*, les études effectuées au laboratoire suggèrent que des mécanismes génétiques pourraient être à l'origine du polymorphisme de ces gènes par des recombinaisons générant de nouveaux allèles. Ils seraient ainsi régulés par des modifications transcriptionnelles, post-transcriptionnelles où les transcrits sont pourraient être soumis à des mécanismes d'épissage alternatif, trans-épissage et épissage aberrant (Roger *et al.* 2008b), et par des modifications post-traductionnelles comme les glycosylations.

Les mécanismes contrôlant l'expression et la régulation de toutes ces molécules diversifiées ne sont pas totalement clairs. La diversification extraordinaire des molécules démontrée chez l'hôte et le parasite ne serait pas uniquement le fruit des mécanismes génétiques. Ainsi nous avons émis l'hypothèse que cette variabilité pouvait trouver des bases au travers d'un mécanisme épigénétique.

1.5.5. Régulation épigénétique des interactions hôte / parasite

Dans le but d'approfondir nos connaissances à propos des mécanismes de diversification, nous sommes partis de notre hypothèse de départ, basée sur le « *Dual inheritance system* ». Non seulement des mécanismes génétiques, mais aussi des mécanismes épigénétiques, contribuent à l'expression du phénotype sous influence de l'environnement (Maynard Smith 1990; Danchin & Wagner 2010).

Des études effectuées dans d'autres modèle hôte / parasite, ont montré que des mécanismes épigénétiques constituent une source de variations phénotypiques qui est mise en place rapidement permettant une coévolution relativement rapide (Gómez-Díaz *et al.* 2012).

Au contact hôte / parasite, le parasite sera exposé au système immunitaire de l'hôte. Ce parasite exprime des antigènes de surface lui permettant de contourner la reconnaissance de l'hôte. Chez certains parasites eucaryotes, les gènes codant les protéines de surface ont évolué d'une façon particulière leur permettant de coder des protéines hautement diversifiées. Ces gènes appartiennent souvent à des familles multigéniques, dont les différents membres sont souvent générés suite à des évènements de duplication (Roger *et al.* 2008b; Verstrepen & Fink 2009). Chez certains parasites, il a été suggéré que des mécanismes épigénétiques sont responsables de contrôler l'activation/inactivation de la transcription de ces gènes (Ralph & Scherf 2005; Perrin *et al.* 2013).

Plasmodium falciparum, eucaryote unicellulaire, l'agent responsable de la malaria, première endémie parasitaire mondiale, possède une famille multigénique qui code pour des antigènes nommés gènes *var*. Le génome de chaque individu possède environ 60 gènes *var*, qui sont localisés majoritairement dans les régions

subtélomériques. Au contact de l'hôte, le plasmodium exprime un seul ou un très faible nombre de gènes et le reste des gènes restent silencieux. Cette transcription monoallélique est sous contrôle épigénétique, régulée par des modifications des histones à l'origine du statut hétérochromatinien (Scherf *et al.* 2008; Verstrepen & Fink 2009).

Un autre exemple qui illustre l'implication des mécanismes épigénétiques dans une interaction hôte / parasite : le cas du parasite unicellulaire *Trypanosoma brucei* responsable de la trypanosomose (maladie de sommeil). Le génome de ce parasite code au moins pour 1250 à 1500 gènes Variant Surface Glycoproteins (VSG) (Berriman *et al.* 2005). A l'instar de plasmodium, ces gènes de surface du parasite sont localisés dans des régions subtélomériques. Ils se caractérisent également par une transcription monoallélique due à un contrôle épigénétique. En effet, c'est un cas d'une expression sélective d'un gène qui dépend de la position de son site de transcription dans le noyau et de son accessibilité aux facteurs de transcription (Navarro *et al.* 2007; Verstrepen & Fink 2009).

I.5.5.1. Epigénétique chez *S. mansoni* et *B. glabrata* : Etat des connaissances

Généralement dans les interactions hôte / parasite, la pression de sélection est élevée. Les deux partenaires, l'hôte *B. glabrata* et son parasite *S. mansoni* évoluent rapidement. Cette coévolution trouve ses validations moléculaires dans l'exemple des molécules hautement diversifiées exprimées par chacun des partenaires, comme nous venons d'en discuter précédemment. D'un point de vue trait de vie, cette coévolution s'illustre par le phénomène de compatibilité de chacune des souches du parasite *S. mansoni* vis-à-vis du mollusque.

Comme chez d'autres systèmes hôte / parasite, est-ce que ces mécanismes épigénétiques peuvent être à l'origine de la variabilité phénotypique observée à la fois au niveau moléculaire et aux différents traits de vie entre *B. glabrata* / *S. mansoni*?

Avant de se lancer dans cette question, il est essentiel de mettre en avant l'état des connaissances au niveau de l'épigénétique chez chacun des deux partenaires.

I.5.5.1.1. Epigénétique chez *S. mansoni*

Bien que les mécanismes épigénétiques moléculaires ne soient pas encore tous clairs chez le parasite, nous avons des éléments et des outils pour étudier les mécanismes épigénétiques chez *S. mansoni*. Ce parasite se caractérise par un génome relativement petit, qui fait 369 Mb (Berriman *et al.* 2009) avec 48% de séquences répétées (Lepesant *et al.* 2012). Son génome est diploïde formé par 7 paires d'autosomes et une paire de chromosomes sexuels (ZZ pour le mâle et ZW pour la femelle) (Berriman *et al.* 2009). La disponibilité du génome séquencé a permis tout d'abord de déterminer la présence des gènes impliqués dans la machinerie des mécanismes épigénétiques.

Méthylation de l'ADN : Le génome de *S. mansoni* code uniquement la DNMT2 parmi les ADN méthyltransférases dont le rôle en tant qu'ADN méthyltransférase est controversée (Goll *et al.* 2006). La modification chimique de l'ADN par méthylation semble faire l'objet d'un débat. En effet, des études montrent que la méthylation de l'ADN est indétectable chez *S. mansoni* (Fantappiè *et al.* 2001; Raddatz *et al.* 2013). Alors que d'autres équipes ont montré que la méthylation de la cytosine existe chez *S. mansoni* et joue un rôle bien déterminé dans le processus de la ponte des œufs (Geyer *et al.* 2011). La méthylation de l'ADN chez *S. mansoni* reste donc une question ouverte. Dans tous les cas, la méthylation de l'ADN chez cet organisme, si elle existe, reste un phénomène minoritaire.

ARN non codants : Récemment, il a été montré que les mécanismes de l'interférence par des petits ARN (RNAi), régulant l'expression génique chez les eucaryotes, sont présents aussi chez le parasite *S. mansoni* (Krautz-Peterson *et al.* 2010; Oliveira *et al.* 2011). Cai et ses collaborateurs ont caractérisé l'ensemble des petits ARNs non codants chez une autre espèce de schistosome, *S. japonicum*, montrant un profil d'expression qui est différent en fonction des différents stades du développement (Cai *et al.* 2013). Cela suggère que les ARN non codants interviendraient dans les mécanismes du développement du parasite. Des travaux récents suggèrent que des miRNA interviennent dans la maturation sexuelle après l'appariement des couples chez *S. japonicum* (Sun *et al.* 2014).

Modification des histones : C'est la base épigénétique la plus étudiée chez le parasite jusqu'à l'heure actuelle. En effet, le génome de *S. mansoni* code toutes les protéines histones (H3, H4, H2A et H2B), ainsi que leurs variants. Non seulement les histones, mais aussi les gènes codants la machinerie enzymatique de leur modification est présente dans le génome du parasite. Ce sont des résultats de recherche par homologie dans la base de données SchistoDB (<http://schistodb.net/schisto/>) des séquences annotées chez *S. mansoni* (Zerlotini *et al.* 2009).

Les membres de mon équipe ont déjà mis au point les techniques permettant d'analyser les modifications des histones par immunoprécipitation de la chromatine (ChIP). Cette technique permet l'étude de la relation entre les modifications des histones et l'expression de certains gènes.

Brièvement, le ChIP consiste tout d'abord à digérer l'ADN entre les nucléosomes à l'aide d'une nucléase micrococcane. L'échantillon digéré est incubé par la suite avec des anticorps qui reconnaissent spécifiquement des marques d'histones. Etant donné que les histones sont des protéines hautement conservées, un bon nombre d'anticorps commerciaux sont disponibles. Une étape de mise au point de la spécificité et de l'efficacité de l'anticorps sur le modèle étudié, est nécessaire (voir l'annexe de la thèse). Différents anticorps ont permis de montrer que les modifications canoniques d'histones sont présentes dans le génome de *S. mansoni*, telles que H3K4me3, H3K9me3, H3K9Ac, H3K27me1 et d'autres (Cosseau *et al.* 2009). Certaines sont considérées comme des marques d'euchromatine telle que H3K9Ac et d'autres comme des marques d'hétérochromatine telle que H3K9me3 (Wang *et al.* 2008).

Afin d'analyser des régions cibles ou des gènes particuliers, nous effectuons un ChIP suivi par de la PCR quantitative (ChIP-qPCR). Alors que pour une analyse globale du génome, le ChIP est suivi d'une étape de séquençage massif (ChIP-Seq) (Cosseau *et al.* 2009). Ces études permettent de corréler le statut chromatinien d'une région bien déterminée avec son état transcriptionnel.

Dans mon équipe, des épigénomes du parasite *S. mansoni* ont été générés pour la première fois par ChIP-Seq (Lepesant *et al.* 2011). Par une étude comparative globale des épigénomes du parasite, nous avons pu distinguer les différences de structure chromatinienne autour des gènes entre les deux différentes souches du parasite *Sm* BRE et *Sm* GH2, me permettant de choisir les meilleurs candidats moléculaires « épigénétiquement variables » pour mes études (voir Chapitre II).

Par ailleurs, par une comparaison globale des épigénomes des parasites stressés avec un anti-helminthe (hycanthone) et non stressés, mes collègues ont démontré que des mécanismes épigénétiques pouvaient être responsables de l'apparition de certains phénotypes de résistance à un anti-helminthe chez *S. mansoni*. En effet, après le traitement des vers parasites par l'hycanthone, il est constaté des changements d'enrichissement de certaines marques épigénétiques au niveau de certains locus. Ceci suggère que des changements de la structure chromatinienne s'opèrent en réponse à un stress chez *S. mansoni* (Roquis *et al.* 2014).

I.5.5.1.2. Epigénétique chez *B. glabrata*

Chez le mollusque *B. glabrata*, les mécanismes épigénétiques ont été très peu décrits avant ma thèse. Les travaux de thèse d'Edwin C. Odoemelam sous la direction de Matty Knight ont suggéré l'importance de la topographie nucléaire d'un locus et l'éventuelle implication des histones modifiées pour cette localisation (Odoemelam 2009). Ils ont montré que suite à une infestation du mollusque par des miracidia de *S. mansoni in vivo*, certains gènes se repositionnent de la périphérie du noyau vers le centre. Ces gènes, notamment le gène de réponse au stress *heat shock protein (HSP 70)*, montrent ainsi une surexpression. Ils suggèrent d'après leur étude que le parasite est responsable d'une réorganisation spacio-épigénétique du génome hôte. Ceci lui permettrait de détourner la machinerie de l'hôte à son profit, en régulant l'expression / surexpression de certains gènes (Knight *et al.* 2011; Arican-Goktas *et al.* 2014).

Mes travaux de thèse ont consisté à initier les premières recherches sur l'épigénétique chez *B. glabrata*. Nous nous sommes focalisés sur la description de porteurs de l'information épigénétique dans le but d'ouvrir des perspectives de recherche liées à

l'implication de l'épigénétique dans la coévolution entre le mollusque *B. glabrata* et le parasite *S. mansoni*.

Nous avons utilisé plusieurs ressources disponibles à la communauté scientifique pour initier l'étude de l'épigénétique chez *B. glabrata*.

- Le génome du mollusque fait 931 Mb, soit 2.5 fois plus grand que celui du parasite *S. mansoni*. Pour le projet « Génome du mollusque *B. glabrata* » un consortium a été établi en 2005 (Raghavan & Knight 2006) et son annotation est actuellement en cours <https://www.vectorbase.org/organisms/biomphalaria-glabrata>.
- Des données transcriptomiques de *B. glabrata* sont disponibles et publiées sur le site (SnailDB).
- Nous disposons au laboratoire des données transcriptomiques globales de différentes souches de *B. glabrata* qui sont en cours d'annotation.

Bien que l'épigénétique du mollusque n'était pas encore étudiée, des études concernant les mécanismes épigénétiques chez d'autres invertébrés, ainsi que leur rôle sont déjà en cours d'étude. La marque la plus étudiée à l'heure actuelle chez les invertébrés, est la méthylation de l'ADN. Cependant, les données restent encore limitées surtout chez les mollusques.

1.5.5.2. Régulation épigénétique des *SmPoMucs* chez *S. mansoni*

Comme mentionné avant, les *SmPoMucs* constituent une famille multigénique qui code les marqueurs clés de compatibilité du parasite avec le mollusque. Ce sont des mucines hautement diversifiées, soumises à différents mécanismes de polymorphisme géniques, modifications post-transcriptionnelles et modifications post-traductionnelles (Roger *et al.* 2008b).

Par ailleurs des travaux menés précédemment dans mon équipe de recherche ont montré qu'après le traitement des miracidia à la Trichostatin A (TSA), la transcription de certains *SmPoMucs* augmente (Perrin *et al.* 2013) et de nouveaux variants sont produits (figure 18) (Cosseau *et al.* 2010). La TSA est un inhibiteur réversible des histones désacétylases HDAC classe I et classe II (Yoshida *et al.* 1995). Utilisé comme un agent épimutagène, la TSA est susceptible de modifier les marques

épigénétiques et par conséquent moduler l'expression des gènes. Chez *S. mansoni*, l'activité enzymatique des HDACs est globalement réduite par la TSA pour tous les stades de développement du parasite (Dubois *et al.* 2009). Ainsi, au travers l'utilisation de cet inhibiteur des histones désacétylases, les travaux antérieurs de mon équipe ont montré que l'expression des *SmPoMucs* semblait être sous contrôle épigénétique, en partie, via des modifications post-traductionnelles d'histones. Il est important de souligner que d'autres travaux ont contribué à montrer l'importance de ces modifications d'histones pour le développement du parasite et sa plasticité phénotypique (Azzi *et al.* 2009).

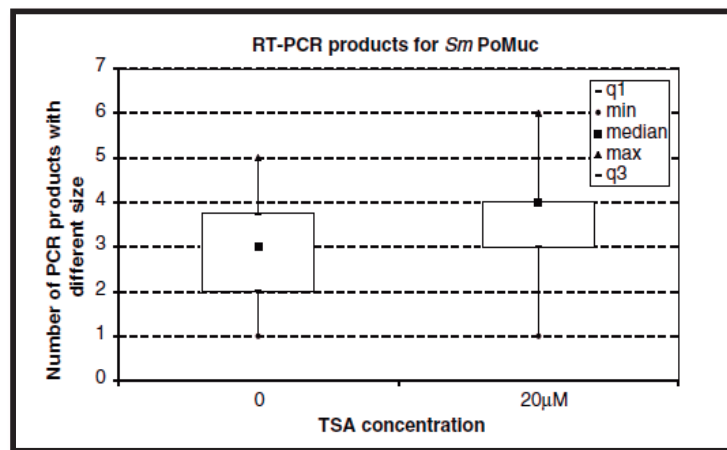


Figure 18 : Augmentation du nombre de transcrits des *SmPoMucs* en fonction du traitement à la TSA. Cette figure montre le nombre de produits obtenus par RT-PCR de différents gènes appartenant à la famille des antigènes du parasite *SmPoMucs* avant et après le traitement à la TSA (Cosseau *et al.* 2010).

Par conclusion, les mécanismes responsables des variations phénotypiques héréditaires constituent actuellement des questions fondamentales dans la biologie évolutive. Traditionnellement, les variations génétiques au niveau de la séquence de l'ADN étaient la seule source des variations du phénotype. Cependant les études actuelles montrent l'existence de différents mécanismes non génétiques tel que l'épigénétique, qui pourraient être à l'origine totale ou partielle de ces variations phénotypiques.

I.6. Objectifs de la thèse

De nouveaux systèmes générateurs de variabilité phénotypique, tel que l'épigénétique, doivent être pris en compte pour expliquer ces exemples d'évolution rapide.

Le but de cette recherche doctorale était de déterminer la contribution aussi bien qualitative que quantitative, des mécanismes épigénétiques dans les variations phénotypiques héritables.

L'interaction *S. mansoni* / *B. glabrata* constitue un bon modèle pour étudier l'évolution adaptative. En effet, c'est une interaction hôte / parasite au cours de laquelle le parasite cherche à s'adapter le mieux possible dans son hôte. A son tour, l'hôte cherche tous les moyens pour résister au parasite. Une meilleure caractérisation de cette capacité d'adaptation à son environnement pourrait conduire à mieux comprendre les bases de l'expansion de cette maladie.

Les principales questions que nous nous sommes posées au cours de ma thèse sont les suivantes :

Chez *S. mansoni* :

Chez le parasite, nous disposons des données et des outils moléculaires qui nous ont permis de poser les questions suivantes :

1. Les modifications épigénétiques seraient-ils des sources des variants phénotypiques?
2. Quel est le poids relatif du « génotype » et de « l'épigénotype » dans l'expression du phénotype ?
3. Ces modifications épigénétiques suivent-ils une hérédité mendélienne?

Chez *B. glabrata* :

Nous avons initié les travaux de caractérisation de l'épigénétique chez cet organisme. Pour cela, nous nous sommes attachés à répondre aux questions suivantes :

1. Quels sont les porteurs d'informations épigénétiques chez *B. glabrata* ?
2. La méthylation de l'ADN existe-t-elle chez *B. glabrata* ? Quel est le taux de 5mC ? 5hmC ? Quel type de méthylation présente cet invertébré ?

Nous disposons au laboratoire de différentes souches et espèces de parasites et de mollusques. Durant ma thèse, j'ai essentiellement travaillé avec deux isolats géographiques différents. Les souches brésiliennes : *Sm* BRE avec son mollusque sympatrique *Bg* BRE, et les souches guadeloupéennes : *Sm* GH2 avec son mollusque *Bg* GUA. La figure 19 illustre les deux différents isolats géographiques, le code couleur (orange pour les souches brésiliennes et verte pour les souches guadeloupéennes). Ce code couleur est utilisé pour le reste des parties de ma thèse.

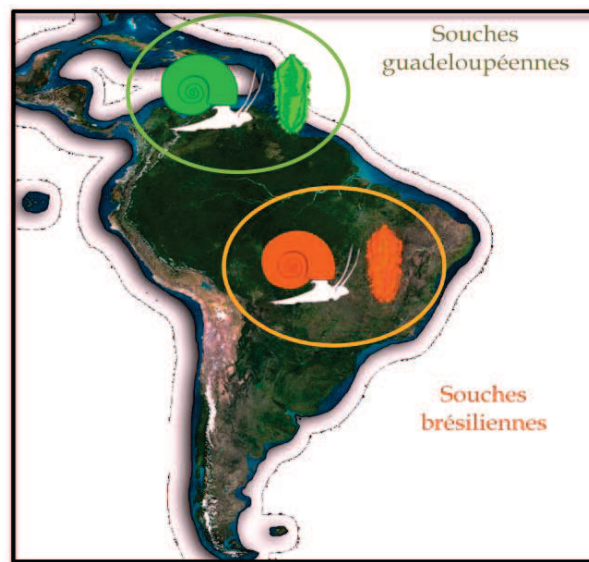


Figure 19 : Les deux différentes souches d'isolats géographiques de *B. glabrata* / *S. mansoni*. Les souches brésiliennes *Bg* BRE / *Sm* BRE en couleur orange et les souches guadeloupéennes *Bg* GUA/ *Sm* GH2 en couleur verte.

Les raisons pour lesquelles ces deux souches ont été choisies :

- (i) Avec ces deux isolats géographiques, nous avons un certain nombre de ressources phénotypiques. En effet, les traits de vie de ces souches en relation avec la compatibilité (prévalence et intensité) ont été très étudiés au laboratoire : *Sm* BRE et *Sm* GH2 se caractérisent par une différence de compatibilité vis-à-vis de la même

souche du mollusque. Les données de prévalence et d'intensité de ces 2 souches constituent donc des **traits de vie** avec lesquels j'ai travaillé pendant ma thèse.

(ii) Les études moléculaires comparatives globales, ont mis en évidence chez ces deux souches des déterminants moléculaires clés pouvant être à l'origine de ce polymorphisme de compatibilité, les ***SmPoMucs***. Nous possédons ainsi un phénotype moléculaire que nous pourrions tenter de lier aux traits de vie.

Finalement, les résultats sont divisés en deux chapitres (chapitre II et chapitre III). Le chapitre II porte sur le rôle de l'épigénétique dans la régulation des *SmPoMucs* (publication N°1) & les bases épigénétiques des variants phénotypiques adaptatifs chez le parasite *Schistosoma mansoni* (publication N°2). Le chapitre III, quant à lui, porte sur la mise en évidence des porteurs de l'information épigénétique chez l'hôte *Biomphalaria glabrata* (publication N°3). Les résultats des travaux publiés et non publiés sont discutés dans le chapitre IV. Les conclusions et les perspectives à l'issue de ce travail sont proposées dans le chapitre V.

Chapitre II

Implication de l'épigénétique dans le succès d'infestation chez le parasite

Schistosoma mansoni

Chapitre II : Implication de l'épigénétique dans le succès d'infestation chez le parasite *Schistosoma mansoni*

Ce chapitre est divisé en deux parties :

Partie 1 : Le rôle de l'épigénétique dans la régulation des *SmPoMucs*.

Partie 2 : Les bases épigénétiques des variants phénotypiques adaptatifs chez le parasite *Schistosoma mansoni*.

L'évolution biologique est faite d'une confrontation permanente entre la variabilité phénotypique et les pressions sélectives. Comme je l'ai mentionné précédemment, les sources de variabilités phénotypiques peuvent être à la fois génétiques et épigénétiques (Pal & Miklos 1999; Klironomos *et al.* 2013). Le but de mon travail a été de comprendre ou de déterminer leur rôle clé dans un contexte de coévolution hôte / parasite sur le modèle *B. glabrata* / *S. mansoni*.

D'une manière générale, cette interaction est dynamique et nécessite la mise en place de variants phénotypiques très diversifiés du côté de l'hôte et de celui du parasite (Frank 2000).

Les variants moléculaires caractérisés par E. Roger et Y. Moné, les *SmPoMucs*, sont des protéines hautement diversifiées aussi bien à l'échelle populationnelle qu'à l'échelle individuelle du parasite (Roger *et al.* 2008b; Moné *et al.* 2010) et constituent de très bons candidats moléculaires pouvant être à l'origine du polymorphisme de compatibilité de l'interaction du parasite avec le mollusque (Mitta *et al.* 2012). Les études menées précédemment au laboratoire ont montré que la TSA (modulateur chimique de l'épigénome) influençait l'expression des *SmPoMucs* (Cosseau *et al.* 2010) et nos hypothèses sur l'implication de l'épigénétique comme source de variabilité phénotypique nous ont conduit à caractériser les bases épigénétiques de la régulation des *SmPoMucs*.

Le chapitre II fait état de ces travaux.

II.1. Partie 1 : Le rôle de l'épigénétique dans la régulation des *SmPoMucs*.

II.1.1. Contexte de la partie 1

A mon arrivée au sein de l'équipe, cette partie 1 était en cours d'étude. Les analyses génétiques, phylogénétiques des promoteurs des *SmPoMucs* ont été effectuées par Cécile Perrin et ses collègues au cours de son Post-doc (2010-2011). Par ailleurs, ces analyses ont été complétées par les résultats de ChIP au niveau des promoteurs de différents groupes de *SmPoMucs*. Ces travaux sont basées sur une étude comparative entre les souches *Sm* BRE et *Sm* GH2. En effet, il est connu qu'il existe une différence d'expression des *SmPoMucs* entre ces souches (figure 16 de l'introduction) et le but de ces travaux était d'étudier si des différences génétiques et/ou épigénétiques pouvaient expliquer ces différences de régulation. Les principaux résultats de cette étude sont :

- Une étude comparative entre les séquences de l'ADN au niveau des promoteurs des *SmPoMucs* de *Sm* BRE et *Sm* GH2 a été menée. Les promoteurs de différents groupes et sous-groupes de *SmPoMucs* (groupe 1, groupe 2 et groupe 3) ont été séquencés pour les deux différentes souches du parasite. Il a été montré ainsi que les différences épigénétiques sont plus importantes que les différences génétiques au niveau des promoteurs. Les différences nucléotidiques entre les deux souches montrent qu'il existe de 0 à 8 substitutions nucléotidiques dans la région promotrice de différents groupes de *SmPoMucs*. Aucune substitution n'a été observée dans la boîte TATA ni dans le TSS des promoteurs entre les deux différentes souches.
- Les régions promotrices séquencées ont été fusionnées à des gènes rapporteurs Enhanced Green Fluorescent Protein (EGFP), et ont permis de montrer que ces régions sont fonctionnelles et suffisantes pour avoir une transcription.
- La structure chromatinienne dans ces mêmes régions a été étudiée par immunoprécipitation de la chromatine (ChIP) en utilisant deux anticorps qui reconnaissent deux marques chromatiniennes différentes. H3K9Ac qui est supposée être une marque d'euchromatine et H3K9me3 qui est connue comme une marque d'hétérochromatine (Berger 2007; Wang *et al.* 2008). Ces deux marques H3K9Ac et

H3K9me3 sont choisies surtout parce qu'elles sont mutuellement exclusives et chaque modification est associée à différents types de chromatine. Le ratio H3K9me3/Ac peut être utilisé pour caractériser l'état chromatinien. Ces analyses ont montré clairement des différences de marques épigénétiques entre les régions promotrices des groupe 1 et groupe 3.1 (r1r2) entre les deux souches du parasite. Un enrichissement plus fort pour la marque H3K9Ac avec un enrichissement plus faible pour la marque H3K9Me3 ont été observés dans la souche *Sm* GH2 que dans la souche *Sm* BRE. Ces résultats sont confirmés par des analyses d'expression de différents groupes de *Sm*PoMucs suite à un traitement des miracidia à la TSA, qui a entraîné une augmentation de la transcription de certains groupes.

Ma contribution dans ce travail :

Je me suis investie dans l'étude de l'expression de ces protéines chez les parasites hybrides des deux différentes souches du parasite *Sm* BRE et *Sm* GH2 (Voir Partie II).

Il est important de noter que dans cette publication les deux souches du parasite sont nommées « C » pour la souche compatible *Sm* BRE et « IC » pour la souche incompatible *Sm* GH2. Par contre, dans les deux publications suivantes, nous avons utilisé la nomenclature en fonction des origines géographiques, pour le parasite : *Sm* BRE et *Sm* GH2 ; pour le mollusque : *Bg* BRE et *Bg* GUA.

II.1.2. Publication N°1

Schistosoma mansoni Mucin Gene (*SmPoMuc*) Expression: Epigenetic Control to Shape Adaptation to a New Host

Cecile Perrin, Julie M. J. Lepesant, Emmanuel Roger, David Duval, Sara Fneich, Virginie Thuillier, Jean-Francois Alliene, Guillaume Mitta, Christoph Grunau, Celine Cosseau.

PLOS Pathogens, 2013, 9, e1003571.



Schistosoma mansoni Mucin Gene (*SmPoMuc*) Expression: Epigenetic Control to Shape Adaptation to a New Host

Cecile Perrin^{1,2}, Julie M. J. Lepesant^{1,2}, Emmanuel Roger³, David Duval^{1,2}, Sara Fneich^{1,2}, Virginie Thuillier^{1,2}, Jean-Francois Alliene^{1,2}, Guillaume Mitta^{1,2}, Christoph Grunau^{1,2}, Celine Cosseau^{1,2*}

1 Université de Perpignan Via Domitia, Perpignan, France, **2** CNRS, UMR 5244, Ecologie et Evolution des Interactions (2EI), Perpignan, France, **3** Center for Infection and Immunity of Lille, Inserm U1019, CNRS UMR 8204, Institut Pasteur de Lille, University Lille Nord de France, Lille, France

Abstract

The digenetic trematode *Schistosoma mansoni* is a human parasite that uses the mollusc *Biomphalaria glabrata* as intermediate host. Specific *S. mansoni* strains can infect efficiently only certain *B. glabrata* strains (compatible strain) while others are incompatible. Strain-specific differences in transcription of a conserved family of polymorphic mucins (*SmPoMucs*) in *S. mansoni* are the principle determinants for this compatibility. In the present study, we investigated the bases of the control of *SmPoMuc* expression that evolved to evade *B. glabrata* diversified antigen recognition molecules. We compared the DNA sequences and chromatin structure of *SmPoMuc* promoters of two *S. mansoni* strains that are either compatible (C) or incompatible (IC) with a reference snail host. We reveal that although sequence differences are observed between active promoter regions of *SmPoMuc* genes, the sequences of the promoters are not diverse and are conserved between IC and C strains, suggesting that genetics alone cannot explain the evolution of compatibility polymorphism. In contrast, promoters carry epigenetic marks that are significantly different between the C and IC strains. Moreover, we show that modifications of the structure of the chromatin of the parasite modify transcription of *SmPoMuc* in the IC strain compared to the C strain and correlate with the presence of additional combinations of *SmPoMuc* transcripts only observed in the IC phenotype. Our results indicate that transcription polymorphism of a gene family that is responsible for an important adaptive trait of the parasite is epigenetically encoded. These strain-specific epigenetic marks are heritable, but can change while the underlying genetic information remains stable. This suggests that epigenetic changes may be important for the early steps in the adaptation of pathogens to new hosts, and might be an initial step in adaptive evolution in general.

Citation: Perrin C, Lepesant MJM, Roger E, Duval D, Fneich S, et al. (2013) *Schistosoma mansoni* Mucin Gene (*SmPoMuc*) Expression: Epigenetic Control to Shape Adaptation to a New Host. PLoS Pathog 9(8): e1003571. doi:10.1371/journal.ppat.1003571

Editor: Matty Knight, George Washington University School of Medicine and Health Sciences, United States of America

Received: February 11, 2013; **Accepted:** June 27, 2013; **Published:** August 29, 2013

Copyright: © 2013 Perrin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The funders are L'Agence nationale de la recherche through the Program Schistophepigen (ANR-07-BLAN-0119-02) and the program EPIGEVOL (ANR-2010-BLAN-1720-01, <http://www.agence-nationale-recherche.fr/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: celine.cosseau@univ-perp.fr

Introduction

The interaction of hosts and parasites is one of the best-studied examples of evolution in a changing environment [1]. Their reciprocal antagonistic co-evolution can be illustrated by an arms race in which host and parasite develop mechanisms to circumvent counter-measures developed by their opponents [2], [3]. Under certain conditions, parasite virulence and host defence can be in equilibrium leading to a phenomenon called compatibility. Compatibility occurs in a host-parasite system when the parasite species is capable of infection and transmission through the host species [4]. The phenomenon that some parasite strains are compatible with certain host strains but not with others (and *vice versa*) is called compatibility polymorphism. This phenomenon was described in the platyhelminth *Schistosoma mansoni* and its intermediate host, the mollusc *Biomphalaria glabrata* [5]. *S. mansoni* is a human parasite whose life cycle is characterised by the passage through two obligatory sequential hosts: the fresh-water snail *B. glabrata* (or dependent on the geographical location

other *Biomphalaria* species) for asexual replication, and humans or rodents as hosts for sexual reproduction [6]. The molecular mechanisms underlying compatibility polymorphism between *S. mansoni* and *B. glabrata* were recently investigated by comparing the proteomes of two *S. mansoni* laboratory strains: one strain that is compatible (the C strain) and one that is incompatible (the IC strain) with the same reference *B. glabrata* strain from Brazil [7]. The study identified *S. mansoni* Polymorphic Mucins (*SmPoMucs*) as key markers for compatibility (see [4] for a recent review). *SmPoMuc* glycoproteins have a mucin-like structure with an N-terminal domain containing a variable number of tandem repeats (VNTR) [8]. *SmPoMuc* proteins are highly polymorphic [8] and interact with the Fibrinogen RELATED Proteins (FREPs) of the mollusc [9]. FREPs are diversified antigen recognition molecules playing a central role in the secondary immune response to digenetic trematodes [10], [11], [12]. The extraordinary level of *SmPoMuc* polymorphism is generated by a complex cascade of mechanisms, a “controlled chaos”, acting at the transcriptional, translational and post-translational level [8]. *SmPoMucs* are

Author Summary

Schistosoma mansoni is a parasitic worm and agent of a disease that causes a considerable economic burden in African and South American countries. The propagation of the parasite requires passage through a freshwater snail of *Biomphalaria* genus. In the field, actually very few snails are infected. This is due to the fact that specific strains of the parasite can infect only specific strains of the snail. Comparative studies have shown that this so-called compatibility is based on the expression of a family of genes that are called *SmPoMucs*. We have shown previously that all parasites strains possess the repertoire of all *SmPoMuc* genes but every strain and even every individual parasite expresses only a subset. These differences could be due to DNA sequence differences in the regions that control gene expression, but here we show that these regions are nearly identical. Instead, the chromatin structure shows strain-specific characteristics. This means that the parasite can adapt to different snail strains simply by changing its chromatin structure and not necessarily the DNA sequence. If this holds true for other parasites, then we have to rethink the way parasite evolution is currently imagined but this also provides a new potential entry point to control the spread of diseases.

encoded by a multigene family with at least 10 members that are organised in 4 clusters on the genome. They recombine frequently and generate new alleles [8]. Each individual miracidium (the larva that infects the mollusc) expresses only a specific subset of *SmPoMuc* genes. The mechanisms controlling this expression polymorphism of *SmPoMucs* remained unclear. Our recent finding that Trichostatin A, a modifier of chromatin structure, influences *SmPoMuc* transcription patterns [13] suggests that epigenetic mechanisms participate in transcription control.

Epigenetic information is information on the status of gene activity that is heritable, for which changes are reversible and that is not based on the DNA sequence [14], [15], [16]. The scientific debate about the reason of the evolution of an epigenetic inheritance system (EIS) in most organisms is intense. Others and we have suggested that EIS provides a basis for modifications in the reaction norms that do not require changes of genotypes [17], [13], resulting in increased phenotypic plasticity at the individual level or increased phenotypic variability at the population level. If EIS influences the capacity to generate different phenotypes, both the better adapted phenotype and the capacity to generate this phenotype will be selected for and carried into the next generation. This hypothesis has been largely validated in the malaria parasite *Plasmodium falciparum* which displays “Clonally Variant Gene Expression” (CVGE) [18]. Genes that show CVGE are present in multicopy, such that individual parasites within an isogenic population express these genes at very different levels, often fully active or completely silenced. Their transcriptional patterns are clonally transmitted to the next generations through asexual multiplication, and stochastic changes of the transcription level occur at low frequency. This bet hedging strategy allows for stochastic generation of phenotypic diversity and can be controlled by epigenetic based events, similar to those described for the *var* gene family. The *var* genes encode the red blood cell surface antigen *P. falciparum* erythrocyte membrane protein 1 (PfEMP-1) and their “CVGE” regulation strategy is responsible for surface antigen variation that ultimately results in immune evasion. In this context, the EIS that leads to “CVGE” allows for

rapid adaptation to the ever-changing vertebrate immune environment. In *S. mansoni* miracidia, we have shown that epigenetic-based events influence the phenotypic plasticity in populations [13] and particularly regulate *SmPoMuc* gene expression. To gain further insight into the precise mechanism of regulation of these genes, in the present study we investigated the genetic and epigenetic changes that occurred during the evolution of the phenotypic compatibility polymorphism in two *S. mansoni* strains. We focused on the sequences of the promoters of active *SmPoMuc* genes and investigated whether there exist differences in the promoter sequences between *S. mansoni* compatible and incompatible strains. Our study reveals that IC and C strains display very little within strain genetic variability, and limited nucleotide differences between promoter sequences of the two strains, but show strong chromatin structure differences. These chromatin structures are heritable throughout the life cycle and transmitted to the next generation, therefore demonstrating that EIS can control a heritable adaptive trait, such as compatibility polymorphism.

Results

Transcription of *SmPoMuc* genes is different in IC and C strains of *S. mansoni*

SmPoMuc genes are classified into 4 groups (Roger et al. 2008) according to their 3' region: group 1 to 4. Group 3 is itself divided into subgroups (3.1, 3.2, 3.3 and 3.4). *SmPoMucs* genes have a 5' region containing a variable number of tandem repeats (exon2), which have been previously called r1 and r2 [8]. r2 exclusively occurs in the group1 and 2 and the intermingled r1–r2 exclusively occurs in the subgroup 3.1, which is present in several copies with either the r1–r2 intermingled repeats or with r1. Due to the very high degree of sequence similarity between the *SmPoMuc* groups, specific transcriptional analyses of the different *SmPoMuc* groups were only possible for groups 1, 2 and 3.1(r1–r2). The transcription levels of these groups were compared between miracidia of the IC and C strains. *SmPoMuc* gene groups 1, 2 and 3.1(r1–r2) are 2.2 to 4.9, 2.5 to 6.7 and 18.6 to 59.7 fold more transcribed in the IC than in the C strain, respectively (fig. 1). The 3.1 subgroup containing

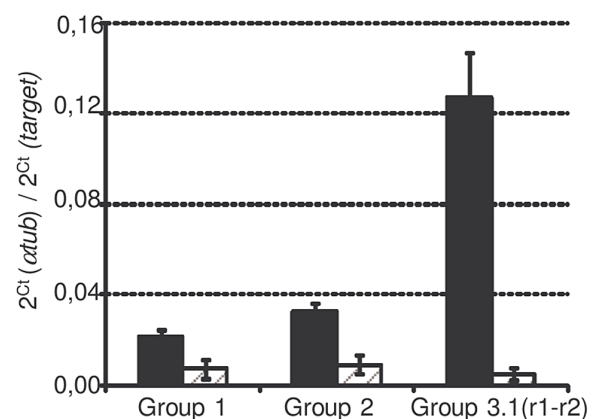


Figure 1. *SmPoMuc* expression of each group in C and IC strains. mRNA were extracted from miracidia pool from the IC (black bars) and C (dashed grey bars) strains and qPCR were performed with primers targeting *SmPoMuc* group 1, group 2, group 3.1(r1–r2). Results represents the mean value of 3 biological repeats, * indicates a p-value below 0.05.

doi:10.1371/journal.ppat.1003571.g001

intermingled r1–r2 repeats is highly transcribed in the IC strain but was practically undetectable in the C strain. This result is consistent with a previous study on individuals of the IC and C strains, which showed that variants containing the r1–r2 combinations are only expressed in the IC strain [8].

The *SmPoMuc* minimal promoter region is located within 1,000 bp upstream of the TSS

To investigate the mechanisms underlying differences of transcription between *SmPoMuc* groups and subgroups, we characterized the minimal promoter region of the *SmPoMuc* genes. We sequenced a region spanning 1.04 to 2.00 kb upstream of the transcriptional start site (TSS) for 4 groups of *SmPoMuc* (Groups 1, 2, 3.1 and 3.1(r1–r2)). We produced a PCR product of a 996 bp of the region of the promoter of the group 3.1(r1–r2) and a PCR product of 1002 bp of the group 3.1 just upstream of the transcriptional start site. Plasmids containing these sequences upstream of a reporter gene (EGFP) were transfected into HeLa cells and fluorescence was observed under a microscope (fig. 2). These experiments showed that these sequences are sufficient to drive the heterologous expression of the reporter gene and contain the minimal promoter sufficient for transcription.

Sequence variations of promoter regions of *SmPoMuc* genes between *S. mansoni* IC and C strains are small

As a first approach to investigate a putative genetic basis for the difference in transcription levels between strains, we investigated

the paralogous and orthologous relationships between the four groups of *SmPoMuc* gene promoters and between the two *S. mansoni* IC and C strains using phylogenetic analysis, reciprocal BLAST dot-plots and comparison of repetitive elements, duplication, recombination events and gene conversions (fig. 3). We annotated the sequences and visualised them by colour-coding of blocks with less than 95% identity (fig. 3). A recombination event was detected using BootScan [19], [20], Maximum Chi Square [21], [22] and Sister Scanning [23] methods in RDP3 and the recombination break points were putatively identified (fig. 3). In both strains we observed one duplication in group 3.1(r1–r2) promoters resulting in an insertion, several insertions/deletions (indels) including one large deletion in group 3.1 promoters and probably a recombination event from the group 2 to group 1 promoter. High similarity to a repeated DNA element was detected in the group 2 promoter; however, it constituted only a small fragment of the complete repeat – 61 bp out of 385 bp of the DIVER2 LTR (*Drosophila*).

The estimated divergence time between the IC and C *S. mansoni* strains is about 400 years [6] and the promoter sequences between the two strains are highly conserved (0.000–0.004 net substitutions per site, Table 1). The number of fixed differences between the two strains varied between 0 in the promoter region of *SmPoMuc* group 2 genes, to 3 in group 3.1, 4 in group 1 and 8 in group 3.1(r1–r2) (Table 1). No substitution was observed in the TATA signal, nor in the TSS regions or in putative regulator binding sites of the promoters between the two strains. *SmPoMuc* promoter sequences were divided into four paralogous sequence groups and sequence differences between strains (orthologous relationships) within groups were much less than the differences observed between groups of the *SmPoMuc* gene family - net substitutions per site varied from 0.000–0.004 within groups of promoter sequences between strains compared to 0.024–0.041 between promoter groups (Table 1). The number of *SmPoMuc* promoter sequence differences between strains was equal to or slightly higher than the number of sequence differences for the promoter of the single copy gene *SmFTZ-F1* [24] which shows no difference between strains (Table 1). Six of 14 microsatellite loci also showed no sequence differences between the two strains (one unique allele). The two strains share the molecular evolution and phylogeny of the promoter region of the four groups of the *SmPoMuc* gene family (fig. 3) – indels, recombination and duplication events. These findings indicate that the divergence between groups of the *SmPoMuc* gene family from a common gene ancestor is ancient and largely predates the time of separation between the IC and C strains.

Low level of promoter nucleotide diversity within the IC and C strains

At this stage of the study we hypothesized that *SmPoMuc* expression differences in C and IC strains could be due to nucleotide differences in the promoter regions of the genes. The sequencing of 1.4 kb of *SmPoMuc* group 1 promoter region for 20 and 18 individuals of the IC and C strains respectively, revealed a very low number of alleles and genotypes (Table 2) – one genotype in the IC strain and 3 genotypes in the C strain. In the C strain, sequence variation was minimal, with the three alleles differing by only one base pair from each other, resulting in insignificant nucleotide diversity (Table 2). All individuals were homozygotes. The IC strain allele of the *SmPoMuc* promoter group 1 differed from the three C strain alleles by four to five base pairs, a sequence divergence of 0.29 to 0.36%. In summary, nucleotide sequence differences between the two strains are surprisingly small.

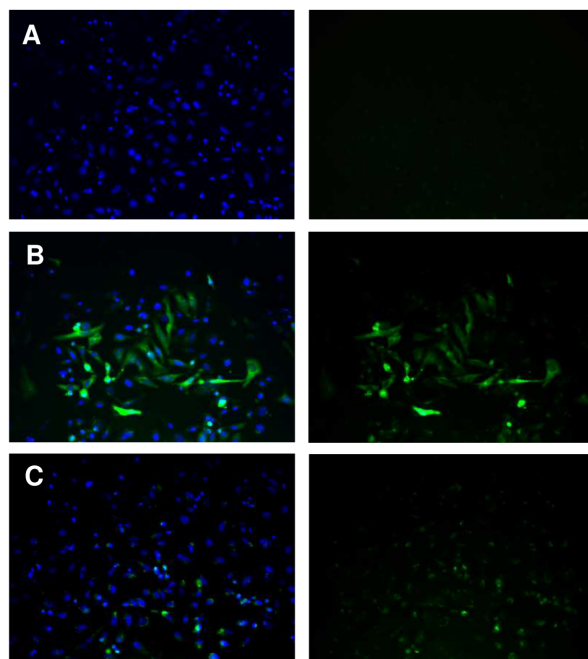


Figure 2. Expression of EGFP under control of the *SmPoMuc* group 3.1(r1–r2) promoter in HeLa cells. HeLa cells were transiently transfected with (A.) promoterless pEGFP1 vector backbone, (B.) pCMV-EGFP or (C.) p*SmPoMuc*-EGFP. Cell nuclei were labelled with DAPI. Visualization of the fluorescence separately (right panel) or overlaid (left panel) is presented. Magnification is $\times 400$. doi:10.1371/journal.ppat.1003571.g002

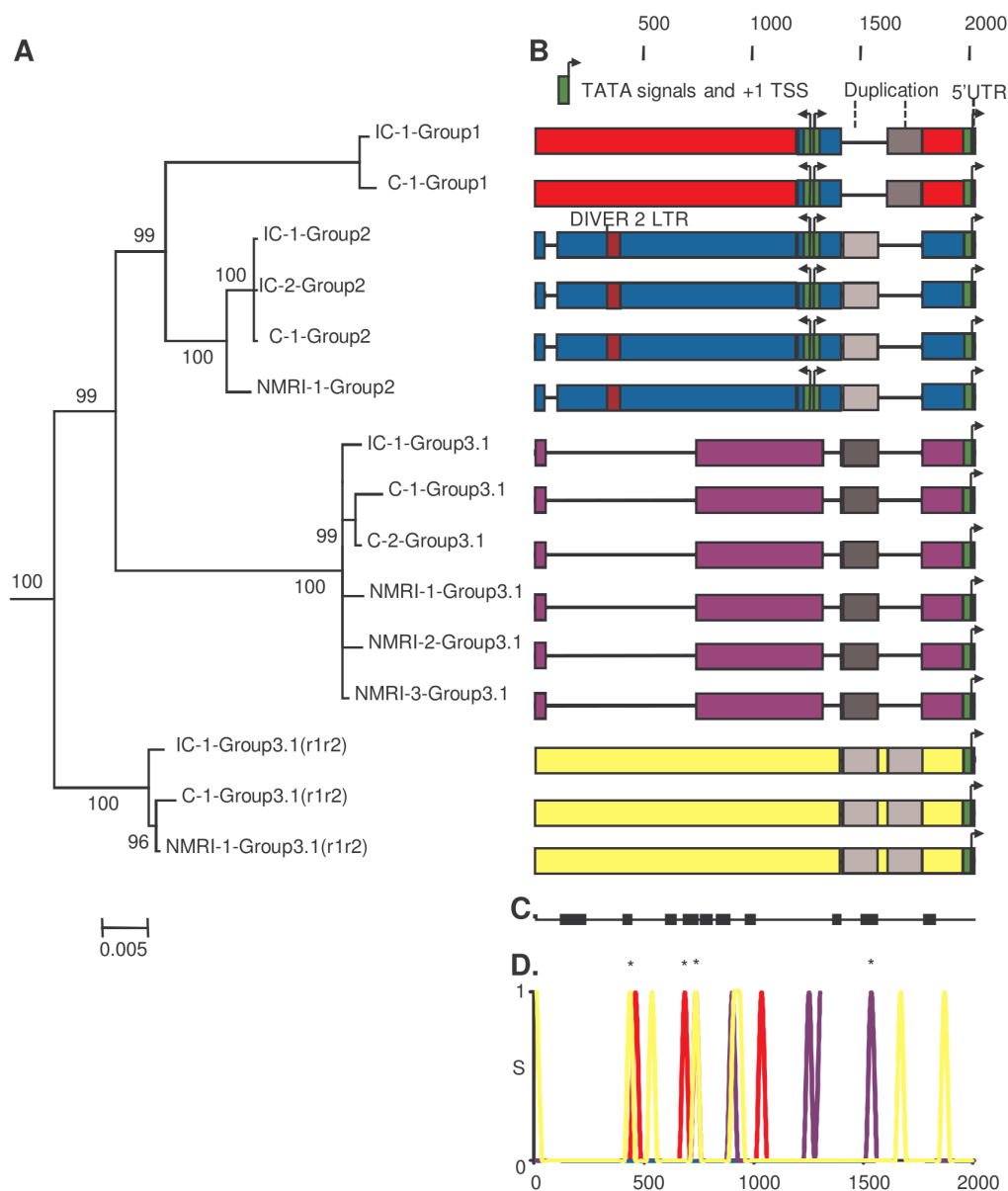


Figure 3. Paralogous and orthologous relationships among *SmPoMuc* promoters between *S. mansoni* IC and C strains. (A.) Bayesian analysis of phylogenetic relationships among *SmPoMuc* promoter sequences with posterior probability values above 70 indicated on associated nodes. (B.) Schematic diagram of aligned *SmPoMuc* promoter sequences corresponding to sequences in panel A. Numbers show the nucleotide position in relation to the transcriptional start site in the alignment. We annotated the sequences by colour-coding blocks of less than 95% identity: Group 1 (red), Group 2 (blue), Group 3.1 (purple) and Group 3.1(r1-r2) (yellow). The 5'UTR were characterised and are represented in orange. TATA signals, here in green, and Transcription Starting Sites (+1 TSS) were predicted using Neural Network Promoter Prediction Tool. Deletions are represented by black lines. A recombination event was detected from Group 2 to Group 1 promoter sequences (in blue in Group 1 sequences). One duplication event resulted in an insertion in Group 3.1(r1-r2) (in greys). Traces of a retrotransposon insertion (DIVER2 – LTR) are present in Group 2 (in burgundy). (C.) Black blocks are the conserved regions among groups of *SmPoMuc* promoters, ignoring the variation within groups of promoters (nucleotide positions as above). (D.) Number of substitutions per site among promoter sequences between strains within *SmPoMuc* groups (S on Y-axis, colour-codes described above), along the sequence alignment (nucleotide positions on X-axis as above). * are the substitutions between the two strains positioned within regions conserved among groups of promoters. The number of substitutions among *SmPoMuc* promoters between the two strains varied from 0 in Group 2 to 8 in Group 3.1(r1-r2). No substitution was observed in TATA signals and TSS sites between the two strains. The sequences have GenBank accession number JQ615951 to JQ615965 (See Table S1 for details). doi:10.1371/journal.ppat.1003571.g003

Table 1. Sequence variation between *Schistosoma mansoni* IC and C strains and between groups of *SmPoMuc* promoter sequences.

Sequences	Seq. length (kb)	N	Nb polym. sites	Nb mutat.	Nb fixed diff. IC vs. C	D_a IC vs. C	D_a group 1 vs. C	D_a group 2 vs. C	D_a group 3.1 vs. C	D_a group 3.1(r1-r2) vs. C
<i>SmPoMuc</i> promoters										
Group 1	1.80	2	4	4	4	0.002	-	0.021	0.031	0.035
Group 2	1.74	3	0	0	0	0.000	-	-	0.040	0.026
Group 3.1	1.04	3	4	4	3	0.003	-	-	-	0.040
Group 3.1(r1-r2)	2.00	2	8	8	8	0.004	-	-	-	-
<i>SmPoMuc</i> introns										
Group 1	7.14	2	12	16	16	0.002	-	0.068	0.031	-
Group 2	7.72	2	32	32	32	0.004	-	-	0.068	-
Group 3.1	7.11	2	43	43	43	0.006	-	-	-	-
<i>SmFTZ-F1</i> promoter										
<i>SmFTZ-F1</i>	1.4	3	0	0	0	0.000	-	-	-	-

Seq. Length, sequence length; N, number of sequences; Nb polym. sites and Nb mutat., number of polymorphic sites and mutations, respectively; Nb fixed diff. IC vs. C, number of nucleotide differences between IC and C strains' promoter sequences; D_a IC vs. C, number of net nucleotide substitutions per site between IC and C strains' promoter sequences; D_a group X vs., number of net nucleotide substitutions per site between groups of promoter sequences.

doi:10.1371/journal.ppat.1003571.t001

SmPoMuc group 1 promoter population sequence difference between IC and C strains is not higher than the average genome-wide difference

Promoter diversity within strain and divergence between strains of *SmPoMuc* group 1 genes were similar to those of 14 microsatellite loci that can be used to reflect genome-wide diversity and divergence [25]. The promoter diversity of *SmPoMuc* group 1 was 0.00 (one allele) in the IC strain compared to 0.22 (3 alleles) in the C strain (Table 2), while expected heterozygosity was 0.000 (one allele) for both strains for 14 microsatellite loci (Data not shown). All individuals were homozygotes. Six out of 14 microsatellite loci showed no divergence between the two strains. At eight microsatellite loci, the IC strain alleles differed from the C strain alleles by one to eleven microsatellite repeats. The promoter region of the single copy *SmFTZ-F1* gene displayed a unique sequence common to the two strains. We estimated extremely high and significant genetic differentiation between the two strains for both *SmPoMuc* group 1 promoter sequences and microsatellite loci using θ , Φ_{ST} and R_{ST} estimators (Table 3). However, we detected almost no heterozygotes and highly significant inbreeding coefficients f in both strains and for both *SmPoMuc* group 1 promoter sequences and the microsatellite loci (Table 3). Therefore the high

values of divergence are likely the result of the bottleneck induced during the care of the life cycle in the laboratory in the two strains as discussed previously [25]. Nonetheless, the distribution of alleles matched the pattern of differentiation as we detected fixed alleles that were different in the two strains. We reasoned that the small genetic differences in the promoter region are simply a by-product of clonality and not the reason for expression differences. We therefore explored an alternative hypothesis, *i.e.* that the expression differences are due to dissimilarity in the epigenetic information.

HDAC inhibitors have an effect on *SmPoMuc* transcription

As the difference in *SmPoMuc* transcription phenotype cannot easily be explained by genetic differences in the promoter region, we investigated the putative implication of epigenetic mechanisms. As a previous study had shown that histone modifications are clearly involved in *S. mansoni* epigenetic mechanisms [13], [26], we tried to influence the epigenotype and phenotype (*SmPoMuc* expression pattern) of *S. mansoni* using trichostatin-A (TSA) that is a specific and reversible inhibitor of class I and II histone deacetylases (HDAC). Treatment with this drug prevents histone deacetylation and is expected to increase the overall acetylation of histones and therefore gene expression [26][27]. The influence of TSA treatment on the transcription of *SmPoMuc* genes (group 1, 2 and 3.1(r1-r2)) of both C and IC strains was tested in miracidia larvae exposed during 4 h to the drug. A Friedman non-parametric test was performed to test the significance of the TSA effect (Figure S1). We observed a statistically significant increase in transcription of groups 1 and 2 after TSA treatment in the IC strain only (p-value = 0.05). This indicates that changes in histone acetylation correlate with increased expression for *SmPoMuc* group 1 and 2 in the IC strain and has no effect in the C strain. Control genes were also tested for their response to TSA in order to determine that its effect was not pleiotropic. No effect of TSA was observed for these genes (GAPDH, Smp_011030, Smp_152710.1, Smp_054160, Smp_158110.1, GST.B, Glyoxalase, data not shown).

Table 2. Sequence diversity of the promoter of *SmPoMuc* group 1 genes for *Schistosoma mansoni* IC and C strains.

	N	A	S	Gen. Div. \pm SD	Nuc. Div. \pm SD	Tajima's D	P
IC	20	1	0	0.00 \pm 0.00	0.000 \pm 0.000	0.000	1.000
C	18	3	2	0.22 \pm 0.12	0.000 \pm 0.000	-1.508	0.048

N, number of *S. mansoni* samples; A, number of alleles; S, number of substitutions; Gen. Div. \pm SD and Nuc. div. \pm SD, expected unbiased gene diversities and nucleotide diversities, respectively. Tajima's D, P, Tajima's D and test of significance of D, respectively.

Sequence length: 1.4 kb.

doi:10.1371/journal.ppat.1003571.t002

Table 3. Genetic differentiation between *S. mansoni* IC and C strains estimated by Φ_{ST} and θ , and inbreeding coefficients f within strains.

	Φ_{ST}/R_{ST}	θ	f in IC	f in C
<i>SmPoMuc</i> promoter group 1	0.974 ⁵	0.898 ⁵	N/A	1.00 ⁵
Microsatellites	six loci N/A (one allele)	six loci N/A (one allele)	fourteen loci N/A	N/A
	eight loci 1.000 ⁵	eight loci 1.000 ⁵		

⁵: significant departure from 0 at $P=0.05$.

doi:10.1371/journal.ppat.1003571.t003

Strain hybrids express both C and IC strain specific *SmPoMuc*s

Since the TSA treatment influences overall histone acetylation, it could not be excluded that the observed effect is an indirect one and that *SmPoMuc* expression control is posttranscriptional and/or posttranslational such as selective RNA or protein degradation. We reasoned that in the offspring of crosses between the IC and C strains transcriptional control would produce an additive pattern of *SmPoMuc* proteins, while control by selective degradation of gene products would produce a subtractive pattern. Western blots show that in miracidia that are produced from crosses between the strains an additive pattern of the C and IC specific bands can be observed (fig. 4). This indicates that regulation operates at the transcriptional and not the post-transcriptional level and further supports the view that chromatin structure plays a role in the generation of specific *SmPoMuc* profiles for each strain.

There are strong chromatin status differences in the *SmPoMuc* promoters between *S. mansoni* C and IC strains

Since all our experiments had delivered results in favour of a difference in chromatin structure of the *SmPoMuc* locus between strains, we decided to investigate the chromatin status in these loci. The occurrence of DNA methylation in *S. mansoni* is currently

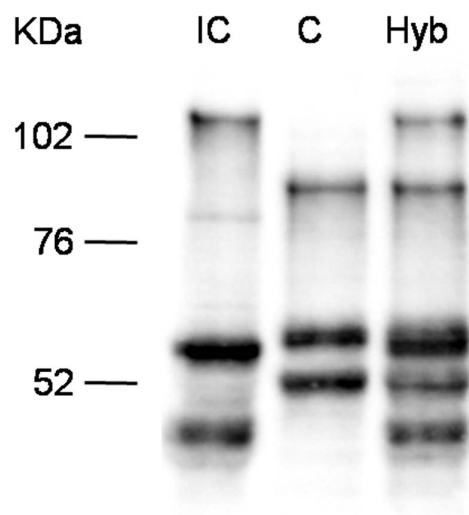


Figure 4. Western blots analysis on *SmPoMuc* proteins from C/IC hybrids and C and IC control strains. Western blot experiments were performed on miracidial protein extracts from C/IC hybrids and C and IC control parental strains.
doi:10.1371/journal.ppat.1003571.g004

debated [28][29]. To test for DNA methylation in the promoter region of *SmPoMuc*s we performed bisulfite genomic sequencing of DNA from miracidia using *in-vitro* methylated DNA as a positive control. We did not detect any methylated cytosine in the target region while 98% of the CpGs of *in-vitro* methylated DNA scored methylation positive. Our results are in line with earlier results showing that DNA methylation is rare from genes in *S. mansoni* [29][28]. We then performed Chromatin ImmunoPrecipitation (ChIP) experiments to check for histone modifications in the promoter regions. Due to the high similarity between the different groups of *SmPoMuc* promoters, ChIP-qPCR (quantitative Polymerase Chain Reaction) analysis was possible only in degenerate regions. Therefore, the chromatin structure analysis was performed on the promoter regions of *SmPoMuc* groups 1, 3.1 and 3.1(r1–r2). ChIP experiments were performed using an antibody that recognised Histone 3 acetylated on lysine 9 (H3K9Ac) and Histone 3 tri-methylated on lysine 4 (H3K4Met3) which are euchromatic marks and an antibody that recognised H3 tri-methylated on lysine 9 (H3K9Met3), which is a heterochromatic mark. Immunoprecipitation with the antibody that targets H3K4Met3 did not show any enrichment in the *SmPoMuc* region tested for either the IC or C strains whereas controls, α Tub (Smp_090120.2) and 28S (Z46503.1) were positive (data not shown). The H3K4Met3 mark is usually very sharp and difficult to localise by target approach. Both *SmPoMuc* group 1 and 3.1(r1–r2) from the IC strain displayed a higher level of H3K9Ac compared to the C strain (fig. 5). Consistent with this result, the C strain displayed a higher level of the heterochromatic mark (H3K9Met3) for group 1 and 3.1(r1–r2). These results have been obtained with several generations of the parasite, demonstrating that the phenotype is transmitted to the next generation.

In the IC strain, epigenetic marks showed differences among *SmPoMuc* groups 1, 3.1 and 3.1(r1–r2) (Figure S2). The promoter of group 3.1(r1–r2) is the most acetylated and the least heterochromatic. This result is consistent with expression analysis after TSA treatment where no effect of TSA was observed for the expression of group 3.1(r1–r2). This absence of an effect of TSA may be explained by the fact that acetylation on this promoter is already saturated and cannot be further increased as previously observed for H4 acetylation in the promoter region of HDAC1 in *S. Mansoni* [26].

The chromatin status in the promoter sequence of *SmPoMuc* groups 1, 3.1 and 3.1(r1–r2) was also investigated in the IC strain in cercaria and adults where *SmPoMuc* genes are not expressed. The level of the heterochromatic and euchromatic marks was the same as in miracidia and this level was maintained through several generations (Figure S3).

Discussion

The host-parasite arms race determines that variability-generating processes are crucial for survival on both sides of the

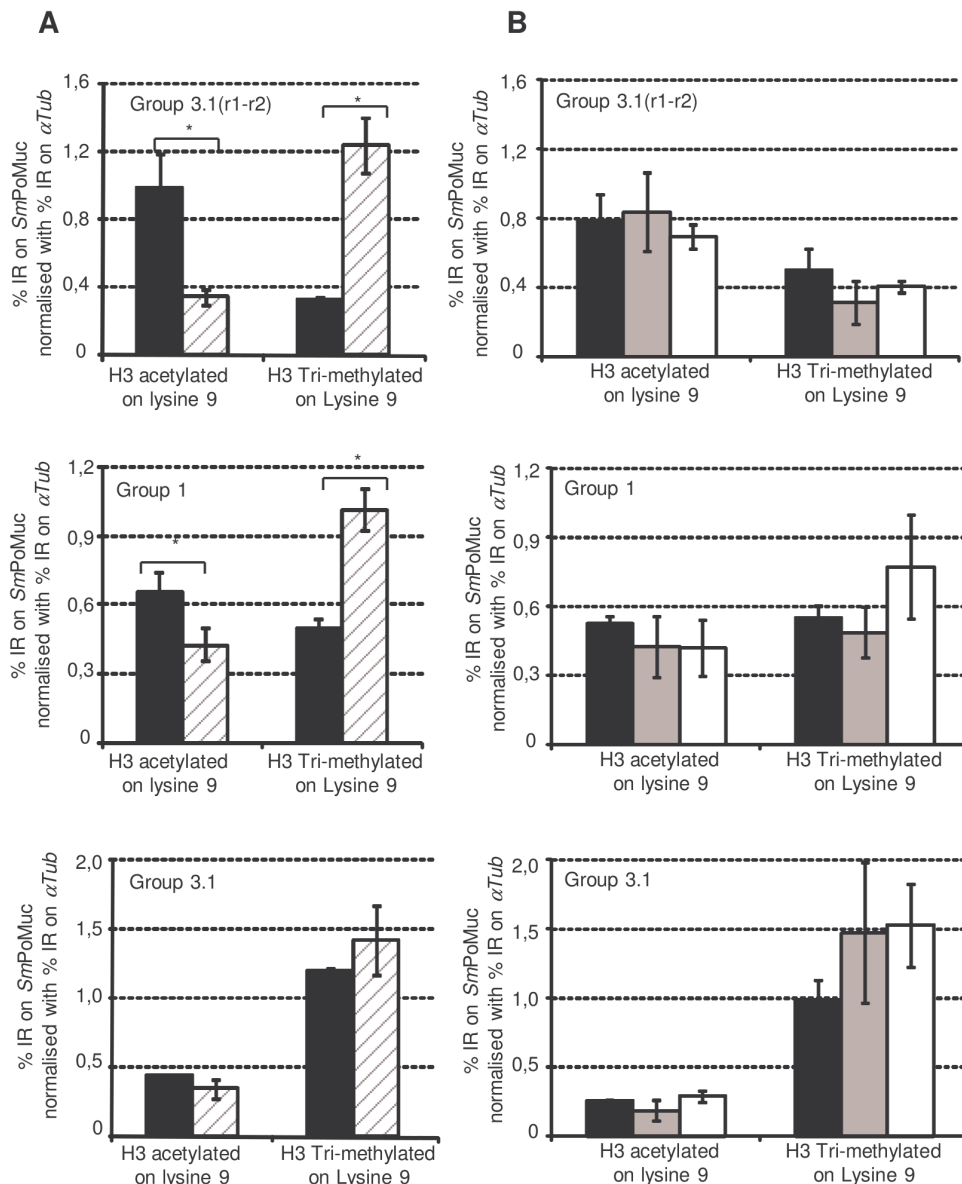


Figure 5. Chromatin immunoprecipitation on *SmPoMuc* promoter regions. Experiments were performed on chromatin isolated from miracidia from both the IC (black bars) and C strains (dashed grey bars) (A.) and on chromatin isolated from IC strain miracidia (Black bars), cercaria (grey bars) and adults (white bars) (B.). ChIP was performed with antibodies against H3 acetylated on lysine 9 and H3 tri-methylated on lysine 9. Immunoprecipitated chromatin was analysed by qPCR using primers that hybridize with specific sequences of *SmPoMuc* group 3.1(r1-r2), group 1 and group 3.1. Results represent the percentage input recovery (%IR) on target genes normalised with %IR of a reference gene (α Tub). Results are the average of 3 biological samples. * indicates a p-value < 0.05 of a student t test.
doi:10.1371/journal.ppat.1003571.g005

interaction (red queen hypothesis, [2]). The mechanisms that are responsible for these (heritable) phenotypic variations are a current and fundamental question in evolutionary biology. Traditionally, random genetic changes are seen as the sole source of phenotypic variation. But the picture is probably more complex: heritable adaptive phenotypic shifts could be partly controlled by epigenetic factors that were underrated until recently [30], [31]. A high rate of heritable epigenetic changes would generate phenotypic variation, which in turn could allow a rapid response to selection pressures [13]; [32]. This could allow for a transient and efficient

response to changes in the environment, and could subsequently be followed by stabilization through genetic changes [33], [34]. Epigenetic modifications affect the transcription status of a gene in a heritable way without changes in the DNA sequence [14], [15], [16] and epigenetic information can be based on a chromatin marking system. Chromatin exists either as a relaxed structure that is permissive to gene expression and is called euchromatin, or as a condensed structure that is typically silent and is called heterochromatin [35]. Therefore, these different chromatin states alter gene expression and, ultimately, influence phenotypic outcomes

without changes to the DNA sequence. The evolutionary implications of epigenetic inheritance systems and their potential link to stress-induced phenotypic variation have been discussed in several models [36], [37], [38], [31], [39], [40], [41] as well as in the specific context of host-pathogen interaction [42].

While it is clear now that induced epigenetic modifications are heritable [43], there are very few reports that show that epigenetic events lead to modification of gene expression profiles, production of new phenotypes and adaptation to the environment [44]. In the present work, we addressed the question of the relative importance of genetic and epigenetic differences between two strains of *S. mansoni* that show clear differences in an ecological important adaptive trait: the capacity to infect their intermediate host. We had previously identified the *SmPoMuc* genes as surface molecular markers important for host compatibility. These markers encode mucins that display an extraordinary level of polymorphism, although they are produced from a relatively small number of very similar genes. As we had shown that nucleotide differences in the coding region could not explain differences in transcription, we focused therefore on the promoter regions in the present work. Our comparative survey of sequence variation in the different groups of *SmPoMuc* gene family from IC and C strains revealed a high level of conservation of the promoter sequences of *SmPoMuc* genes between the two strains. The molecular evolution of *SmPoMuc* promoters was uniform between all strains analysed, IC, C and NMRI. The sequence differences between the IC, C and NMRI strains within each group of *SmPoMuc* promoter were small, and the number of substitutions between the IC and C strains was equal or slightly higher than in the monomorphic single-copy gene *SmFTZ-F1* and consistent with sequence differences at 14 microsatellite loci. To assess whether substitutions between the two strains could have an effect on transcription, we searched for functional regions of the active promoters. None of the substitutions between the IC and C strains occurred in the TATA signal, putative transcription factor binding sites or TSS regions. The nucleotide differences between the two strains consisted of zero in group 2 to eight substitutions in group 3.1(r1–r2), resulting in net nucleotide substitutions per site similar or lower than the ones observed in presumably neutral *SmPoMuc* introns (Table 2). At the population level, our analysis of *SmPoMuc* group 1 promoters in the IC and C strains revealed very low allelic and nucleotide variability within strain and high allele frequency differences between the IC and C strains due to fixed substitutions. All individuals were homozygotes at *SmPoMuc* group 1 promoter, similarly to the genotypes at 14 microsatellite loci, suggesting that *S. mansoni* strains present genome-wide homozygosity. Both strains are characterised by a high significant inbreeding coefficient, resulting from high clonality in the two strains [25], which may have arisen because of the bottleneck due to the strain maintenance in laboratory conditions. Despite the lack of diversity within strains, alleles fixed in each strain for the *SmPoMuc* group 1 promoter and nine microsatellites were different, resulting in high genetic differentiation between the two strains as estimated by F_{ST} . This contrasted with the promoter of the single-copy gene *SmFTZ-F1* and six microsatellite loci, which displayed a unique sequence common to the two strains.

In summary, our analysis of the genetic information shows that (i) both strains are genetically monomorphic, including the *SmPoMuc* promoter regions, (ii) both strains are different in terms of alleles, i.e. they do not share the same alleles, but (iii) these alleles are similar or display low number of base substitutions (outside functional regions). It could be argued that the small nucleotide differences observed between the two strains are sufficient to provoke modulation of histone modification. Such a

leverage effect of SNPs cannot be excluded but has so far not been observed in heavily studied models such as human, *Drosophila melanogaster* and *Arabidopsis thaliana*. It could also be the case that strain-specific loci exist that regulate the chromatin structure of the *SmPoMuc* genes *in trans* or *in cis* (upstream of the minimal functional promoter). However, previous work has compared the proteomes of both C and IC strains [7] and did not pinpoint any major regulators that may be responsible for such a phenotype. In view of these results, we argue that genetic differences between sequences within each group of *SmPoMuc* promoters were unlikely to solely dictate the high level of variation in *SmPoMuc* transcription and compatibility polymorphism phenotypes.

We therefore further investigated the epigenetic basis for such phenotypes. TSA treatment was used to study the impact of overall acetylation status of histones on miracidia larvae where *SmPoMuc* is expressed. This drug is known to be a specific histone deacetylase (HDAC) inhibitor and has been previously shown to influence phenotypic traits in *S. mansoni* [13]. A dose dependant effect of TSA was observed for *SmPoMuc* expression (all groups taken together) in the IC strain whereas no effect was observed in the C strain. This result suggests that the acetylation status of histones in the promoter sequences is differentially regulated between the IC and C strains. HDACs seem to play a more prominent role in regulating the acetylation level in the IC strain that allowed us to pinpoint a TSA effect in this strain. More specifically, we report a TSA effect on groups 1 and 2 of the IC strain whereas no effect is observed for group 3.1(r1–r2) for which acetylation is the strongest. This also suggests that a differential regulation by HDAC exists between the *SmPoMuc* groups in the same strain. Further support for regulation on transcriptional level comes from a crossing experiment in which strain hybrids were produced. Western blots show that in the hybrids, both the C-specific and the IC-specific *SmPoMuc*s are expressed. One could hypothesize that production of *SmPoMuc* variants is due to post-transcriptional strain-specific regulation. In this scenario all genes would be expressed, but the gene products would be processed in a strain-specific form. In the hybrids, in which the hypothetical post-transcriptional regulation pathway for both strains is present, we should have seen a diminution of non-IC and the non-C *SmPoMuc* forms. This was not the case. In summary, all lines of evidence point towards a chromatin-based regulation of *SmPoMuc* expression.

The chromatin configuration was further investigated by ChIP analysis using antibody that recognises heterochromatic and euchromatic marks. ChIP results clearly demonstrate that different epigenetic marks occur on the *SmPoMuc* promoter of group 1 and group 3.1(r1–r2) between the IC and C strains likely resulting in a different chromatin configuration. On these loci, chromatin is indeed more enriched in H3 acetylated on lysine 9 in the IC compared to the C strain and less enriched in the opposite mark, H3 trimethylated on lysine 9. Therefore, the local chromatin structures differ between the two strains for groups 1 and 3.1(r1–r2) and are consistent with expression data as stronger acetylation correlates with enhanced expression. Importantly, H3K9Me3 and H3K9Ac marks are maintained through the cercarial and adult stages at which the genes are not expressed. This persistence of the chromatin mark throughout other stages of the *S. mansoni* life cycle is a crucial result as this is a necessary condition for the epigenetic mechanism to act as a heritable trait. Similarly, several CVGE genes of *P. falciparum* that display a bistable chromatin state to regulate their expression in the intraerythrocytic stages have been shown to maintain their epigenetic marks during trophozoite and schizont stages, the other asexual stages at which these genes are not expressed [45].

It is now established that the phenotype is not only a product of genetic processes, but expression of an ensemble that is composed of genetic and epigenetic components. Others and we have proposed that this additional system allows for rapid adaptive evolution without necessarily changing the genotype initially. A theoretical framework for this model was provided by Pal and Miklos (1999) [17], and more recently by Klironomos, Berg and Collins (personal communication). Essentially, these authors propose that a higher rate of random changes in epigenetic marks compared to genetic mutations transmitted from one generation to the next in a population generates increased phenotypic variations that can be selected for if the environment changes. In this sense, epigenetic modifications provide a source of rapid and reversible phenotypic variation and are therefore expected to be major players in the context of host-pathogen interaction where selection pressures are strong and evolution is fast [42], [18]. In this context, epigenetic based events to generate variability of surface antigens of parasites perfectly matched to this theory. For example, VSP diversification of *Giardia sp.* likely occurs by epigenetic mechanisms involving the histone acetylation status [46] and/or RNAi [47]. Chromatin remodeling proteins and histone modifications have been shown to play a role in VSG expression site silencing [48] and *Plasmodium* Var diversification is orchestrated by multiple epigenetic factors including monoallelic transcription at separate spatial domains at the nuclear periphery, differential histone marks on otherwise identical *var* genes, and *var* silencing mediated by telomeric heterochromatin [49]. On the host side, genetic and epigenetic crosstalks have been previously demonstrated in the generation of a high level of polymorphism of the receptors of the adaptive immune system [50], [51]. Therefore, all these variability generating mechanisms are examples of local adaptation to an ever-changing environment where epigenetic based events are used to rapidly produce new phenotypes and potentially induce rapid evolutionary change of genes that are under pressure. In our work, we show that two population of *S. mansoni* with distinct phenotypic traits, in particular their compatibility with a reference host, show low nucleotide differences in both coding sequence and promoters of *SmPoMuc* but high epigenetic differences in the promoter regions. Both parasite populations are in a situation where the fitness value of genetically encoded phenotypes has not changed significantly, but epigenetic variations have produced phenotypic variants that are adapted to different environments (compatible hosts).

While we have compared only South American strains, our observations suggest a scenario for the adaptation of *S. mansoni* to the new world host: in the 15th–16th century the ancestral strain of contemporary strains IC and C migrated via the slave trade from Africa to the West Indian Islands and the South American continent, respectively [6]. There, they had to adapt to a new intermediate host. The initial bottleneck resulting from the migration of only a limited number of parasites and the expected strong selective pressure acting on both genetic and epigenetic variants of the key-molecules for compatibility with the new snail hosts, *SmPoMucs*, may have significantly reduced genetic and epigenetic variation in the newly formed laboratory IC and C strains compared to the ancestral strain. Now, it is likely that epigenetic variation retained from the ancestral strain and the higher rate of occurrence of epigenetic changes in subsequent generations, rather than the strain genetic variation, enabled the parasite to adapt rapidly to their host and new environment. A conundrum with the “epigenetic mutation system first” hypothesis is that epigenetic information concerns the transcriptional activity of a gene but not its coding potential, in other words, a gene can be switched on and off by the surrounding chromatin but the

resulting protein cannot be changed. Loss of function of genes can easily be imagined through an epigenetic mechanism, but for gain of function a complex inhibitor-based mechanism would be necessary. The classical Ohno hypothesis of gene duplications as way to provide material for evolution [52] could deliver a solution. Rodin and Riggs have shown that duplicated genes have a tendency to be heterochromatic [53]. It is interesting to note that the *SmPoMuc* proteins, essential for host compatibility, are encoded by duplicated genes. Our analysis shows that the duplication events predate the IC/C separation and occurred in the strain’s common ancestor, *i.e.* gene duplication was not a result of divergence of the two strains. We postulate that *SmPoMuc* duplicated genes provide an additional system for phenotypic variation. Duplicated genes are randomly modulated in their relative transcriptional activity through chromatin structure changes as evidenced by our current and previous results [13], resulting in new combinations of expressed *SmPoMuc* genes and subsequent increased phenotypic variation. If the parasite encounters new intermediate hosts, the probability for the phenotypes to match is increased, thus allowing for adaptive evolution.

Therefore, our work shows that in a gene family that codes for an adaptive phenotypic trait, epigenetic changes are more important than genetic changes. This finding provides support for theoretical models of adaptive evolution in which epimutations occur more rapidly than mutations.

Materials and Methods

Ethics statement

The French Ministère de l’Agriculture et de la Pêche and French Ministère de l’Education Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848) for the experimenters. Housing, breeding and animal care followed the national ethical requirements.

Culture of *Schistosoma mansoni*

A compatible strain (C) (Brazilian strain), an incompatible *S. mansoni* strain (IC) (Guadeloupean strain), the reference NMRI *S. mansoni* strain (Puerto Rican strain) and a reference mollusc strain (*B. glabrata* BRE isolated from Brazil) were used in this study. For initial breeding, each strain was maintained in its sympatric (compatible) *B. glabrata* strain, and in hamsters (*Mesocricetus auratus*) as described previously [54]. Adult worms and miracidia were obtained as described previously [8].

Generation of strain hybrids and Western blot

Individual *B. glabrata* snails were infested with a single miracidium to obtain cercarial clonal populations. Subsequently the sex of the cercariae was determined as described previously [55]. Strain hybrids of *S. mansoni* were produced by infection of mice or hamster with 300 cercariae: 200 males from a clonal cercarial population combined with 100 females from another clonal cercarial population. Different combinations of parental cercariae of the IC and C strains were used, thus generating worm couples in which the male is C and the female is IC or *vice versa*. Eggs were recovered from infected (3 to 6) mice (*Mus musculus*) 12 weeks post-infection. Livers were collected and homogenized, and eggs were filtered and washed. Miracidia were allowed to hatch in spring water and were concentrated by sedimentation on ice for 15 minutes.

1000 Miracidia were incubated in 350 μ l UTCD buffer (ultrapure urea 8 M, Tris 40 mM, DTT 65 mM, CHAPS 4%), two hours at room temperature. The extract was cleared by centrifugation for 30 minutes at 1500 g, and the supernatant was collected. Total proteins (5 μ g per sample) were separated by 10% SDS-PAGE gel electrophoresis before being blotted on a nitrocellulose membrane (Trans-Blot turbo, Bio-Rad). The membrane was blocked with 5% non-fat dry milk in TBST (TBS buffer containing 0.05% tween 20) one hour at room temperature, and incubated with the primary antibody “anti-*SmPoMuc*” diluted 1/500 in TBST for 90 minutes at room temperature. This rabbit polyclonal antibody was produced according to standard procedures and was shown to recognise all the *SmPoMuc* groups [9]. Then, the membrane was incubated with secondary antibody (peroxidase conjugated, purified anti-rabbit IgG) diluted 1/5000 in TBST for 1 hour. After washing 3 times for 10 minutes in TBST, the detection was carried out using the ECL reagents and the ChemiDoc MP Imaging system – BioRad).

PCR screening for promoters of *SmPoMuc* genes, cloning and sequencing

We searched for sequences of promoter regions of *SmPoMuc* genes in the genomic database of the *S. mansoni* NMRI strain (assembly version 3.1) using BLAST searches. Contigs matching to *SmPoMuc* genes were assembled with the Sequencher software (Gene Codes Corporation) to recover the sequences of the promoter regions of the genes. From the BLAST search and manual assemblage of relevant contigs, scaffolds of promoter regions were constructed for the different *SmPoMuc* genes in groups 1–4. Primers were designed on these contigs to amplify the promoter regions of the different *SmPoMuc* genes in the C and IC strains of *S. mansoni*. The DNA templates to generate PCR products were either genomic DNA (C and IC strains), a BAC library (NMRI strain) or a phage library (IC strain). Genomic DNA was extracted from adult worms as described previously [8]. The production of the phage library is described below. Promoter regions were amplified using the Advantage 2 PCR Enzyme System (Clontech) (Table S1 for primer sequences, amplified fragment lengths and sources of DNA). PCR products were either cloned into pCR-XL-TOPO (TOPO TA Cloning kit for sequencing, Invitrogen) and plasmid DNA was purified using the Wizard Plus SV Miniprep DNA purification system (Promega), or sequenced directly. We sent PCR amplicons or plasmids containing the promoter regions to GATC (GATC Biotech, Germany) for cycle sequencing in both directions and performed primer walking up to 2.0 kb upstream of the transcription start sites (TSS) of *SmPoMuc* genes (for primer sequences see Table S2). We checked trace data and aligned nucleotide sequences manually using the BioEdit software. We scanned the promoter sequences for putative regulator binding sites using the web based interface Program NSITE (Softberry Inc.) (<http://linux1.softberry.com/berry.phtml?topic=nsite&group=programs&subgroup=promoter>).

Production and screening of a phage lambda library of IC genomic DNA

The presence of multiple copies of some *SmPoMuc* genes sometimes prevented the amplification of a single copy and assembly of a gene with its corresponding promoter. To address this problem, we constructed a phage library of the IC strain using the Lambda Fix II vector system from Stratagene. The expected size of inserts was 15 to 23 kb corresponding to the size range of *SmPoMuc* genes (10–30 kb). Details of the construction of the

phage library and screening are available at <http://methdb.univ-perp.fr/epievo/>. Genome coverage of the library was four fold. The library was screened for *SmPoMuc* genes using as a probe UR1, a highly conserved intronic sequence spanning the region between two repeat units of the *SmPoMuc* genes [8]. The probe was labeled with the DIG High Prime DNA Labeling and Detection Starter Kit II using Random primed DNA labeling with digoxigenin-dUTP, alkali-labile and chemiluminescence with CSPD (Roche). Screening was performed according to the manufacturer's instructions. Secondary and tertiary screening rounds were performed with the same probe to isolate individual phage clones. Phages that scored positive for *SmPoMuc* repeat units were screened by PCR using a combination of diagnostic primers for each group of *SmPoMuc* genes (Table S2) with the Advantage 2 PCR Enzyme System (Clontech). Selected phages were subsequently purified and used as templates to PCR amplify *SmPoMuc* group 3.1(r1–r2) as described in the section “PCR screening for promoters of *SmPoMuc* genes, cloning and sequencing”.

Sequence variation of promoter regions of *SmPoMuc* genes between *S. mansoni* IC and C strains

Sequence annotation and promoter prediction. The 5'UTR and ORF were previously characterised using 5'RACE-PCR experiments [56]. The core promoter including a TATA box and the TSS was predicted using Neural Network Promoter Prediction Tool (http://www.fruitfly.org/seq_tools/promoter.html) [57]. We identified repetitive elements in the promoter region sequences using the CENSOR software [58]. We searched for duplications, recombinations and gene conversions using dot plots among sequences and the programs RDP3 [20]. *SmPoMuc* promoter sequences were annotated using CLC Sequence Viewer v6.5.1 (CLC Bio 2011). We colour-coded paralogous sequence blocks, portions of repetitive elements, duplications and recombination to visualise the evolution of paralogous and orthologous *SmPoMuc* promoter sequences. The number of substitutions per site for pairwise comparisons and searched for conserved regions was calculated with DnaSPv4.50.3 [59].

Phylogenetic analysis. We performed Bayesian phylogenetic analyses using MrBayes 3.2.0 [60]. We sampled across the substitution model space in the Bayesian Markov Chain Monte Carlo (MCMC) itself [61]. The model selected was the HKY model. Insertion/deletion (indel) events were coded as binary characters (presence/absence) and included as a separate binary data partition in the analysis [60]. We ran the MCMC for 120,000 generations, trees being sampled every 100 generations. This allowed the final average standard deviations of split frequencies to reach below 0.01 and the potential scale reduction factors (PSRF) for all parameters to be close to 1, indicating that the runs had converged onto the stationary distribution. The first 1,000 trees were discarded as burn-in to compute the consensus tree. We repeated the analyses three times to ensure the posterior probabilities were stable. Trees were rooted with a sequence of the promoter sequence of the *SmPoMuc* pseudogene group 4.

Sequence variation and gene diversity

We used DnaSP to characterise promoter sequence variation within and between groups of *SmPoMuc* promoter sequences as the number of polymorphic sites, number of mutations between strains, net number of substitutions per site between strains and between groups of *SmPoMuc* promoter sequences.

Sequence variation of the promoter region of a single copy gene, *SmFTZ-F1*, between *S. mansoni* IC and C strains

We amplified and sequenced the promoter region of the *SmFTZ-F1* gene. This gene encodes the nuclear receptor fushi tarazu-factor 1alpha and its promoter has been fully characterised [24] in 1 and 2 individuals of *S. mansoni* strains IC and C, respectively, from genomic DNA with primers *Smftzf1-F* (5'-ATGAGATGTTTCTGAGCAATGGC-3') and *Smftzf1-R* (5'-TCTTCTCGTAGCTGAATCTGACC-3') using the Advantage 2 PCR Enzyme System (Clontech). PCR amplicons were then sequenced and analysed for sequence variation and gene diversity as described above.

Heterologous expression of promoter regions of *SmPoMuc* genes

Cell culture and transfection. HeLa cells were maintained in Dulbecco's modified Eagle's medium (DMEM) and 10% fetal calf serum (FCS) containing an antibiotic/antimycotic mixture (penicillin 100 units/ml, streptomycin 0.1 mg/ml, amphotericin B 0.25 µg/ml; Sigma) at 37°C. Transfections were performed on Lab-Tek chamber slides (0.8 cm²/wells) with 250 ng of DNA using jetPRIME according to the manufacturer's instruction (Polyplus transfection). Briefly, 20,000 cells were seeded per well in 350 µl of cell growth medium 24 h prior to transfection. 250 ng of plasmid DNA diluted into 25 µl jetPRIME buffer were incubated with 1 µl jetPRIME transfectant for 10 min at room temperature. The transfection mix was added directly to the cells. After 72 h, we washed HeLa cells with PBS and fixed in -20°C methanol for 5 min. Cells were washed twice with PBS and counterstained with DAPI (100 mg/ml) for 10 sec and mounted with fluorescent mounting medium (Dako). Fluorescence was observed with a Zeiss Axioskop2 (Zeiss) using a camera Leica DC350FX coupled to imaging software (Leica FW4000).

SmPoMuc promoter construction

We amplified 996 kb of the *SmPoMuc* group 3.1(r1-r2) promoter and 1002 kb of the *SmPoMuc* group 3.1 promoters. These sequences are located just upstream of the transcriptional start site and have been amplified from the IC strain. These sequences were amplified using primers containing *SacI* and *BamHI* restriction sites (Table S2). The PCR product was gel-purified (Wizard SV gel and Clean-Up system, Qiagen), digested with both restriction enzymes and cloned into a *SacI* and *BamHI* digested pEGFP-1 reporter vector with T4 DNA ligase (New England Biolabs). The construct was verified by sequencing both DNA strands. Plasmids pEGFP-1 and pCMV-EGFP driving EGFP expression, under the control of the CMV-promoter, were used as negative and positive controls in the transfection assay.

Sequence variation of promoter regions of *SmPoMuc* group 1 gene between *S. mansoni* IC and C strains at the population level

A 3.3 kb region of the *SmPoMuc* group 1 gene promoter region was amplified using primers *SmpomucpromGP3.1.f2* and *BR2* (Table S1) in individuals of each of *S. mansoni* IC and C strain. The PCR products span from 1.8 kb upstream of the TSS to the first repeat unit of the *SmPoMuc* gene and cover the promoter region. 1.4 kb of the promoter region was sequenced for 20 and 18 individuals of the IC and C strains, respectively, by primer walking (Table S2). We used Arlequin 3.1 to characterise *SmPoMuc* group 1 promoter diversity within the two strains as the expected unbiased gene diversity, the nucleotide diversity, corrected for

sample size and incorporating nucleotide information [62]. We tested for sequence variation between the two strains using population comparisons and differentiation in Arlequin 3.1. Estimations incorporated Tamura-Nei distances between sequences and allele frequencies (Nei's Φ -estimator of F_{ST}). The significance of genetic differentiation was tested by permuting the alleles among all samples 2,000 times. We also estimated the inbreeding coefficient in each strain using f and genetic differentiation between the two strains using F_{ST} estimator θ ([63], incorporating allele frequencies only). Inbreeding coefficients and genetic differentiation for departure from the null hypothesis ($f=0$, $\theta=0$) were tested using 2,000 permutations in GENETIX 4.05 [64].

Allelic variation of 14 microsatellite loci between *S. mansoni* IC and C strains at the population level

Nineteen individuals of each of the IC and C strains were genotyped using 14 microsatellite loci [25]. We estimated genetic diversity of microsatellite loci as the mean number of alleles per locus (A) and observed and expected unbiased heterozygosities (H_O and \hat{H} respectively) under the assumption of Hardy-Weinberg equilibrium [62]. We estimated the inbreeding coefficient f in each strain, genetic differentiation between the two strains R_{ST} estimator [65], [66] and the F_{ST} estimator θ as above.

Trichostatin-A treatment, mRNA extraction, cDNA synthesis and transcription analysis

Trichostatin-A (TSA) (invivoGen met-tsa-5) was dissolved in ethanol to 20 mM and added to the 1000 IC or C miracidia pool at 20 µM and 200 µM during 4 h. We had shown previously the effect of TSA at these concentrations on development, morphology, mobility and gene expression without any cytotoxicity for the larvae [13], [27]. To the untreated control, an equal volume of ethanol was added (mock treatment). After 4 h, metamorphosis arrest was observed for larvae treated with TSA at 200 µM as expected for a positive effect with this drug [27]. Miracidia were then spun down at 12,000 g during 5 min and suspended in 100 µl of lysis buffer (Dynabeads mRNA DIRECT Micro kit, Dynal Biotech) in RNase-free tubes and stored at -80°C. Messenger RNAs were extracted using the Dynabeads mRNA isolation Kit according to the manufacturer's instructions. mRNA poly-A residues were eluted from the surface of the paramagnetic beads by a final denaturation step of 10 min at 75°C in 20 µl of Tris-HCl 10 mM. cDNA synthesis was carried out using 10 µl of mRNA in a final volume of 20 µl according to manufacturer's instructions (0.5 mM dNTPs, 0.01 mM DTT, 1× first strand buffer, 2 U RNase out, 10 U SuperScript II RT (Invitrogen) during 50 min at 42°C). After reverse transcription, the cDNAs were purified with the PCR clean-up system (Promega) and eluted into 100 µl 10 mM Tris/HCl (pH 7.5).

Specific primers for qPCR from groups 1, 2 and 3.1(r1-r2) were designed based on sequence alignment performed on cDNA variant representative of each group (Table S2). Their specificity was tested using as template a plasmid in which a cDNA variant of group 1, 2 or 3.1(r1-r2) was cloned. Group 4 genes contain a STOP codon in exon 8 of the gDNA sequence and their cDNA has never been detected. Therefore, transcripts of the group 4 genes were not targeted in this study. Other subgroups were not studied as it was not possible to design specific primers to amplify them. qPCR amplifications were performed as described below. Results were normalised with the α Tub gene. The $2^{-\Delta C_t}$ value was calculated. Statistical tests were performed on at least 3 different biological samples.

Chromatin status of *SmPoMuc* promoters by ChIP-qPCR

Native chromatin immunoprecipitation was performed as described before [67]. Briefly, antibodies against histone isoforms were used to precipitate chromatin in miracidia from IC and C strains (Table S3). DNA was extracted from the precipitated complex and analysed by qPCR using specific primers of *SmPoMuc* groups 1, 3.1 and 3.1(r1–r2). Primers specifically targeting these genes were designed based on sequence alignment of *SmPoMuc* promoter sequences (Table S2). We tested their specificity using as templates plasmids with promoters of group 1, 3.1 or 3.1(r1–r2). It was not possible to design primer sets that would hybridize specifically to the promoter sequences of the other groups or subgroups because conservation in the sequences resulted in cross-amplification between these groups. The amount of target DNA recovered in the immunoprecipitated fraction was quantified by calculating the percent input recovery (% IR) normalised with the percent input recovery obtained with a reference locus (α Tub) as previously described [67].

Chromatin status of *SmPoMuc* promoter region by bisulfite treatment

Bisulfite genomic sequencing was carried out as described in [68] on gDNA extracted from miracidia from the NMRI strain. Amplification was performed using primers BS.IC-1-Group1/1111-1715.48f GATATGTTTTTAAGAAGTAGAAAAGAATA-TT, BS.IC-1-Group1/1111-1715.508r ATAAAAATTTTACACACCCTACTC and BS.IC-1-Group3.1/421-952.29f AT-TGTTTTTTTTTAATTTTAGATATGTTTTTA and two rounds of PCR. 1 μ l of each PCR products were cloned into the TOPO TA vector (Invitrogen) and sequenced. *In-vitro* methylation with M.SssI (NEB) was done as recommended by the supplier. A total of 20 sequences (7 M.SssI treated positive controls and 13 target miracidial gDNA) were aligned with the genomic sequence from GenBank (Bioedit) to visualise the sites of methylated cytosine.

qPCR analysis

qPCR amplifications were performed with 2.5 μ l of immunoprecipitated DNA or cDNA in a final volume of 10 μ l on a LightCycler® 480 II Real Time instrument (1.5 μ l H₂O, 0.5 μ M of each primer, 5 μ l of master mix). The following protocol was used: denaturation, 95°C for 10 minutes; amplification and quantification (40 times): 95°C for 10 seconds, 60°C for 10 seconds, 72°C for 20 seconds; melting curve, 65–97°C with a heating rate of 0.11°C/s and continuous fluorescence measurement, and a cooling step to 40°C. For each reaction, the cycle threshold (Ct) was determined using the “2nd derivative” method of the LightCycler® 480 Software release 1.5. PCR reactions were performed in duplicate and the mean value of Ct was calculated. Correct melting curves were checked using the T_m calling method of the LightCycler® 480 Software release 1.5. The amplification of a unique band was verified by electrophoresis separation through a 2% agarose gel for each qPCR product.

GenBank accession numbers

JQ615951–JQ615966.

Supporting Information

Figure S1 Expression of each *SmPoMuc* groups in C and IC strains and TSA effect. mRNA were extracted from

miracidia pool from the IC (Panel A) and C (panel B) strain and qPCR were performed with primers targeting *SmPoMuc* group 1, group 2, group 3.1(r1–r2). The results of 3 experiments are represented on each graph (Experiment 1: Black bars, Experiment 2: Dark grey bars, Experiment 3: pale grey bars). A Friedman non-parametrical test was performed to test the significance of the increase of expression after TSA was added. The p value of the Friedman test is indicated on each graph.

(TIF)

Figure S2 Immunoprecipitation of miracidia chromatin: Comparison of the chromatin state of the different group within a strain.

ChIP experiments were performed on chromatin isolated from miracidia from both the IC and C strain with antibodies that target H3 acetylated on lysine 9 and H3 tri-methylated on lysine 9. Immunoprecipitated chromatin was analysed by qPCR using primers that target specific sequences of *SmPoMuc* group 3.1(r1–r2), 1 and 3.1. Results represent the percentage input recovery (%IR) normalised with %IR from a reference gene (α Tub). Results are the average of 3 biological repeats. All p value from t-test that compare the results obtained with group 3.1(r1–r2) and group 1, group 3.1(r1–r2) and group 3.1, group1 and group 3.1 are below 0.05 in the IC strain for both antibodies.

(TIF)

Figure S3 Immunoprecipitation of chromatin from miracidia, cercaria and adults over 3 generations.

ChIP was performed on chromatin isolated from IC strain miracidia (Black bars), cercaria (grey bars) and adults (white bars). ChIP was performed with antibodies against H3 acetylated on lysine 9 (panel A) and H3 tri-methylated on lysine 9 (H3K9Met3). Immunoprecipitated chromatin was analysed by qPCR using primers that hybridize with specific sequences of *SmPoMuc* group 3.1(r1–r2), group 3.1 and group 1. Results represent the percentage input recovery (%IR) on target gene normalised with % IR of a reference gene (α Tub) obtained on 3 generations (G1, G2, G3). Results are the average of 2 technical repeats.

(TIF)

Table S1 Origin of the sequences used for phylogenetic analysis of fig. 3.

(DOCX)

Table S2 Primers used in this study.

(DOCX)

Table S3 Antibodies used in this study.

(DOCX)

Acknowledgments

The authors are indebted to Bernard Dejean and Anne Rognon for providing valuable technical support. The authors acknowledge Dr Jérôme Boissier for his contribution to statistical analysis and Ray Pierce for critical reading of this manuscript.

Author Contributions

Conceived and designed the experiments: CC CG GM. Performed the experiments: CP JMJL ER DD SF VT JFA CC. Analyzed the data: CP ER CC. Contributed reagents/materials/analysis tools: CC JFA. Wrote the paper: CC CG CP GM.

References

- Mackinnon MJ, Marsh K (2010) The selection landscape of malaria parasites. *Science* 328(5980): 866–871.
- Van Valen L (1974) Molecular evolution as predicted by natural selection. *J Mol Evol* 3: 89–101.
- Jemmelly NY, Niang M, Preiser PR (2010) Small variant surface antigens and *Plasmodium* evasion of immunity. *Future Microbiol* 5(4): 663–682.
- Mitta G, Adema CM, Gourbal B, Loker ES, Theron A (2012) Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Dev Comp Immunol* 37(1): 1–8.
- Theron A, Coustau C (2005) Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol* 79(3): 187–191.
- Morgan JA, Dejong RJ, Adeoye GO, Ansa ED, Barbosa CS, et al. (2005) Origin and diversification of the human parasite *Schistosoma mansoni*. *Mol Ecol* 14(12): 3889–3902.
- Roger E, Mitta G, Mone Y, Bouchut A, Rognon A, et al. (2008) Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: new candidates identified by a global comparative proteomics approach. *Mol Biochem Parasitol* 157(2): 205–216.
- Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, et al. (2008) Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). *PLoS Negl Trop Dis* 2(11): e330.
- Mone Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, et al. (2010) A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS Negl Trop Dis* 4: e813.
- Adema CM, Hertel LA, Miller RD, Loker ES (1997) A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci U S A* 94(16): 8691–8696.
- Hanington PC, Forsy MA, Dragoo JW, Zhang SM, Adema CM, et al. (2010) Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. *Proc Natl Acad Sci U S A* 107(49): 21087–21092.
- Zhang SM, Zeng Y, Loker ES (2008) Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. *Innate Immun* 14(3): 175–189.
- Cosseau C, Azzi H, Rognon A, Boissier J, Gourièrre S, et al. (2010) Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* – a possible kick-off for adaptive host/parasite evolution. *Oikos* 119: 669–678.
- Umlauf D, Fraser P, Nagano T (2008) The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. *Biol Chem* 389(4): 323–331.
- Dillon N (2008) The impact of gene location in the nucleus on transcriptional regulation. *Dev Cell* 15(2): 182–186.
- Lee JS, Smith E, Shilatfard A (2010) The language of histone crosstalk. *Cell* 142(5): 682–685.
- Pal C, Miklos I (1999) Epigenetic inheritance, genetic assimilation and speciation. *J Theor Biol* 200(1): 19–37.
- Cortes A, Crowley VM, Vaquero A, Voss TS (2012) A view on the role of epigenetics in the biology of malaria parasites. *PLoS Pathog* 8(12): e1002943.
- Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 11(11): 1423–1425.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26(19): 2462–2463.
- Smith J (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34(126–129).
- Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 98(24): 13757–13762.
- Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16(7): 573–582.
- De Mendonca RL, Bouton D, Bertin B, Escriva H, Noel C, et al. (2002) A functionally conserved member of the FTZ-F1 nuclear receptor family from *Schistosoma mansoni*. *Eur J Biochem* 269(22): 5700–5711.
- Bech N, Beltran S, Portela J, Rognon A, Allienne JF, et al. (2010) Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory. *Infect Genet Evol* 10(7): 1039–1045.
- Dubois F, Caby S, Oger F, Cosseau C, Capron M, et al. (2009) Histone deacetylase inhibitors induce apoptosis, histone hyperacetylation and up-regulation of gene transcription in *Schistosoma mansoni*. *Mol Biochem Parasitol* 168(1): 7–15.
- Azzi A, Cosseau C, Grunau C (2009) *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp Parasitol* 121(3): 288–291.
- Geyer KK, Rodriguez Lopez CM, Chalmers IW, Munshi SE, Truscott M, et al. (2011) Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nat Commun* 2: 424.
- Raddatz G, Guzzardo PM, Olova N, Fantappie MR, Rampp M, et al. (2013) Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc Natl Acad Sci U S A* 110(21): 8627–8631.
- Jablonka E, Lamb M (2005) Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. MIT Press, Cambridge.
- Bosdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. *Ecol Lett* 11(2): 106–115.
- Danchin E, Charmantier A, Champagne FA, Mesoudi A, Pujol B, et al. (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet* 12(7): 475–486.
- Jablonka E, Lamb MJ, Avital E (1998) ‘Lamarckian’ mechanisms in darwinian evolution. *Trends Ecol Evol* 13(5): 206–210.
- Pigliucci M, Murren CJ, Schlichting CD (2006) Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* 209(Pt 12): 2362–2367.
- Luijsterburg MS, White MF, van Driel R, Dame RT (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* 43(6): 393–418.
- Rapp RA, Wendel JF (2005) Epigenetics and plant evolution. *New Phytol* 168(1): 81–91.
- Grant-Downton RT, Dickinson HG (2006) Epigenetics and its implications for plant biology 2. The ‘epigenetic epiphany’: epigenetics, evolution and beyond. *Ann Bot* 97(1): 11–27.
- Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. *Nat Rev Genet* 7(5): 395–401.
- Bosdorf O, Zhang Y (2011) A truly ecological epigenetics study. *Mol Ecol* 20(8): 1572–1574.
- Boyko A, Kovalchuk I (2008) Epigenetic control of plant stress response. *Environ Mol Mutagen* 49(1): 61–72.
- Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol* 84(2): 131–176.
- Gomez-Diaz E, Jorda M, Peinado MA, Rivero A (2012) Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS Pathog* 8(11): e1003007.
- Verhoeven KJ, Van Dijk PJ, Biere A (2010) Changes in genomic methylation patterns during the formation of triploid asexual dandelion lineages. *Mol Ecol* 19(2): 315–324.
- Uchida S, Hara K, Kobayashi A, Otsuki K, Yamagata H, et al. (2011) Epigenetic status of Gdnf in the ventral striatum determines susceptibility and adaptation to daily stressful events. *Neuron* 69(2): 359–372.
- Crowley VM, Rovira-Graells N, Ribas de Pouplana L, Cortes A (2011) Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. *Mol Microbiol* 80(2): 391–406.
- Kulakova L, Singer SM, Conrad J, Nash TE (2006) Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol* 61(6): 1533–1542.
- Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, et al. (2008) Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 456(7223): 750–754.
- Rudenko G (2011) African trypanosomes: the genome and adaptations for immune evasion. *Essays Biochem* 51: 47–62.
- Scherf A, Lopez-Rubio JJ, Riviere L (2008) Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol* 62: 445–470.
- Osipovich O, Oltz EM (2010) Regulation of antigen receptor gene assembly by genetic-epigenetic crosstalk. *Semin Immunol* 22(6): 313–322.
- Bergman Y, Cedar H (2010) Epigenetic control of recombination in the immune system. *Semin Immunol* 22(6): 323–329.
- Ohno S, Wolf U, Atkin NB (1968) Evolution from fish to mammals by gene duplication. *Hereditas* 59(1): 169–187.
- Rodin SN, Parkhomchuk DV, Riggs AD (2005) Epigenetic changes and repositioning determine the evolutionary fate of duplicated genes. *Biochemistry (Mosc)* 70(5): 559–567.
- Theron A, Pages JR, Rognon A (1997) *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol* 85(1): 1–9.
- Portela J, Grunau C, Cosseau C, Beltran S, Dantec C, et al. (2010) Whole-genome in-silico subtractive hybridization (WISH)—using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics* 11: 387.
- Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, et al. (2008) Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol* 157(2): 217–227.
- Reese MG (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem* 26(1): 51–56.
- Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7: 474.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19(18): 2496–2497.

60. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3): 539–542.
61. Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol* 21(6): 1123–1133.
62. Nei (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.
63. Cockerham CC, Weir BS (1984) Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* 40(1): 157–164.
64. Belkhir K, Dawson KJ, Bonhomme F (2006) A comparison of rarefaction and bayesian methods for predicting the allelic richness of future samples on the basis of currently available samples. *J Hered* 97(5): 483–492.
65. Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* 142(4): 1357–1362.
66. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139(1): 457–462.
67. Cosseau C, Azzi A, Smith K, Freitag M, Mitta G, et al. (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Mol Biochem Parasitol* 166(1): 70–76.
68. Grunau C, Clark SJ, Rosenthal A (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* 29(13): E65–65.

Schistosoma mansoni Mucin Gene (*SmPoMuc*) Expression: Epigenetic Control to Shape Adaptation to a New Host

Cecile Perrin, Julie M. J. Lepesant, Emmanuel Roger, David Duval, **Sara Fneich**, Virginie Thuillier, Jean-Francois Alliene, Guillaume Mitta, Christoph Grunau, Celine Cosseau.

Les données supplémentaires :

Table S1: Origin of the sequences used for phylogenetic analysis of fig. 3.

Genbank Accession Number	Target	DNA template	Sequence length	Primer Name and Sequence to generate PCR products	Amplicon size
Group 1					
JQ615951	IC-1	gDNA	1.8kb	SmpomucpromGP3.1.f2: 5'-GCCTGCATAAAATAGGGACCG-3'	3.3kb
JQ615952	C-1	gDNA	1.8kb	BR2: 5'-AACTCACCTGTGGGTTTGTCTG-3'	
Group 2					
JQ615953	IC-1	Phage library ^b	1.74kb	Smpomuc2f3: 5'-TGTTGGATGTTTATATCTATGC-3'	2.4kb
JQ615954	IC-2	gDNA	1.74kb	exon1R: 5'-AGAGAATAATTTTCTTGTTTCATTCTTC-3'	
JQ615955	C-1	gDNA	1.74kb		
JQ615956	NMRI-1		1.74kb	Sequence obtained from assembled sequences BACs 41B11, 62F12 ^a	
Group 3.1					
JQ615957	IC-1	gDNA	1.04kb	smpomuc2f2: GGAATAGTCTAGTGATAACG-3'	9.0kb
JQ615958	C-1	gDNA	1.04kb	exon1R: AGAGAATAATTTTCTTGTTTCATTCTTC-3'	
JQ615959	C-2	gDNA	1.04kb		
JQ615960	NMRI-1	BAC47P6 ^a	1.04kb		
JQ615961	NMRI-2	BAC47P6 ^a	1.04kb		
JQ615962	NMRI-3		1.04kb	Sequence obtained from database assembly version 3.1	
Group 3.1 (r1-r2)					
JQ615963	IC-1	gDNA	2.00kb	exon2f : 5'-TTCTTAGCACTACCCAAAGATGAAC-3' exon1R: 5'-AGAGAATAATTTTCTTGTTTCATTCTTC-3'	9.0kb
JQ615964	C-1	gDNA	2.00kb	Prom groupe4 spe2.for : 5'-TTTCATCATTGTGCGGTGTT-3' prom univ3.rev : 5'-GCCTAATTTTAATTCACAATCTACTCA-3'	3.0kb
JQ615965	NMRI-1		2.00kb	Sequence obtained from database assembly version 3.1	

^a The BAC library was previously described by Roger et al. 2008b.

^b The phage library was constructed in this work and is described in the materials and methods section.

Table S2: Primers used in this study

Table 1. *SmPoMuc* primers used for sequencing and primer-walking.

Target	Name	Primer sequence
Promoter group 1	smpomucpromGP3.1.f2	5'-GCCTGCATAAATAGGGACCG-3'
	BR2	5'-AACTCACCTGTGGGTTTGTCTG-3'
	smpomucprom1.r1	5'-TCACATCCGTGACACTTTCAG-3'
	smpomucprom.r1	5'-TGGACACACGGTAAATAAAACTG-3'
	smpomucprom.r12	5'-TCGGGTGGCATTTCATCACGG-3'
	smpomucprom1.r2	5'-CTGAAATCCCGGAAGAACTG-3'
Promoter group 2	smpomuc2.f3	5'-TGTTGGATGTTTATATCTATGC-3'
	exon1R	5'-AGAGAATAATTTCTTGTTCATTCTTC-3'
	smpomucprom.f1	5'-CTACGAGTAGGTGGCTGCAA-3'
	smpomucprom4.r1	5'-AATGTGCACGAAACATGGAA-3'
Promoter group 3.1	smpomuc2.f2	5'-GGAATAGTCTAGTGATAACG-3'
	exon1R	5'-AGAGAATAATTTCTTGTTCATTCTTC-3'
	Smpomucprom.f1	5'-CTACGAGTAGGTGGCTGCAA-3'
	smpomucprom.r12	5'-TCGGGTGGCATTTCATCACGG-3'
	smpomucprom.r3	5'-TTTATATTCGTGATTTCCTGTTC-3'
Promoter group 3.1(r1-r2)	exon2f	5'-TTCTTAGCACTACCCAAAGATGAAC-3'
	exon1R	5'-AGAGAATAATTTCTTGTTCATTCTTC-3'
	smpomucprom.r1	5'-TGGACACACGGTAAATAAAACTG-3'
	smpomucprom.r2	5'-TGGTGCAAATGAAAAGCGTA-3'
	smpomucprom4.r1	5'-AATGTGCACGAAACATGGAA-3'

Table 2. Primers used for phage DNA screening

	Name	Primer sequence	Discrimination criteria
Primer set 1	repetB.F2	5'-TCTCACATTTTCAGGTGACCTC-3'	Group 1 expected size : 1150bp
	Intron3/4R	5'-CACATGCATAGCTAATGTGGTAATG-3'	Group 2 expected size : 1727bp Group 3 and 4: no amplification expected
Primer set 2	repetG.F2	5'-GCTCTCACATTTTCAGATGACTAT-3'	Group 3.1 expected size: 1136bp

	Intron3/4R	5'- CACATGCATAGCTAATGTGGTAATG -3'	Group 1, 2, 4, 3.3, 3.4: no amplification expected
Primer set 3	URnew	5'-GTGGCCATGTAACAAACGCACC- 3'	Group 4 expected size: 1150bp
	Intron3/4R	5'- CACATGCATAGCTAATGTGGTAATG -3'	Group 1, 2, 3: no amplification expected
Primer set 4	repetB.F2	5'-TCTCACATTTTCAGGTGACCTC-3'	Specific for the presence of combinatory event
	repetG.R2	5'-AACTCACCTGTTGGTTCGCTC-3'	

Table 3. Primers used in this study for transcription and ChIP analysis by qPCR

Target	Name	Primer sequence	size (pb)	Tm	Primer efficiency	Specificity
Primer sets for expression analysis by qPCR						
cDNA group 1	<i>SmPoMuc1.E8/9.F</i>	5'- CGAAAGTGCTTACATC GCTG-3'	208	68°C	1.93	All group 1
	<i>SmPoMuc1.E12.R</i>	5'- CGAATAGGCTTCTTTAT GTTG-3'				
cDNA group 2	<i>SmPoMuc2.E4.F</i>	5'- CCAGATCCACAACAGC ACC-3'	260	68°C	2	group 2 (exclude truncated group 2)
	<i>SmPoMuc2.E8.R</i>	5'- ACTTTCGACGTCATCAA GAAG-3'				
cDNA group 3.1(r1-r2)	URG/B.F	5'- CGAACCAACAGGTGAC CTCGC-3'	ND	65°C	2	Specific to subgroup 3.1 with combinatorial event of tandem repeats ^a

	<i>SmPoMuc3</i> E2/E3.R	5'- CTGGAAGTGTGGTTCG CTC-3'				
Primer sets for ChIP analysis by qPCR						
Promo ter group 3.1(r1r 2)	prom-group3.1(r1- r2).for	5'- ATGATTTTACGAGAGGT TCAGT-3'	101	60°C	2	Specific to promoter of subgroup 3.1 with combinatorial event of tandem repeats ^b
	prom-group3.1(r1- r2).rev	5'- GGTGCACTATTCAGCCT AGTCA-3'				
Promo ter group 1	prom-group1- spe4.for	5'- TCTCTACTGAAAGTGTC ACGGA-3'	130	60°C	1.96	
	prom-group1- spe4.rev	5'- TGAGTGAAGGTTTTGCA GCC-3'				
Promo ter group 3.1	prom-group3.1- spe2.for	5'ATACGAGTCAGTATT AGTAAATG-3'	100	60°C	1.91	
	prom-group3.1- spe2.rev	5'- TTTATCGATCGTAACTT TGGAC-3'				

^a Previous studies have shown that *SmPoMuc* genes contain 3 kinds of tandem repeats in their 5' region. Two of them, r1 and r1', display only 1 nucleotide and 1 amino acid difference (Roger et al. 2008b). The third repeat r2 is highly divergent. The sub-group 3.1 is represented by several copies in the genome. One of these copies (and only in this sub-group 3.1) contains the combination of r1 and r2 type of repeats in their 5' region in both the C and IC strain but expression of the

intermingled repeat units is a specific feature of the IC strain (Roger et al. 2008b). In this study, primer sets designed to target the group 3.1 are specific to this particular *SmPoMuc* cDNA that contains the intermingled repeat units.

^b The promoter sequence of this group 3.1 was identified as the promoter of the sub-group 3.1 that contains the intermingled repeat units. This was demonstrated by PCR analysis on a phage of the phage lambda library that contains the sub-group 3.1 (r1-r2) as an insert.

Table 4. Primers used to amplify promoter sequence for heterologous expression analysis

	Name	Primer sequence	Size (bp)
<i>SmPoMuc</i> group 3.1 JQ615957	prom-group3.1exp.for (position 6 on JQ615957)	5'-CCCTTTgagctcATGGAGCAGTGAGTTA-3'	1002
	prom-group3.1exp.rev (position 1007 on JQ615957)	5'-TATggatccTGGCTTTATATATATTTTAC-3'	
<i>SmPoMuc</i> group3.1(r1-r2) JQ615963	prom-group3.1(r1-r2)exp.for (position 975 on refseq JQ615963)	5'-AATTCgagctcACATGGTTTATATACGTATTGTCTTC-3'	996
	prom-group3.1exp.rev (position 1970 on JQ615963)	5'-TATggatccTGGCTTTATATATATTTTAC-3'	

Table S3: Antibodies used for N-ChIP.

Antibody	Host	Product	Lot	Saturating quantity used for N-ChIP
H3K9Ac	Rabbit	Upstate, 07352	DAM1576933	8µl
H3K9Met3	Rabbit	Abcam, Ab8898	733951	4µl
H3K4Met3	Rabbit	Upstate, 04745	JBC1355340	4µl

II.2. Partie 2 : Les bases épigénétiques des variants phénotypiques adaptatifs chez le parasite *Schistosoma mansoni*.

II.2.1. Contexte de la partie 2

L'objectif principal de cette partie c'était de déterminer le poids relatif de chacune des composantes génétique et épigénétique dans l'expression du phénotype chez le parasite *S. mansoni*. Ceci sous-entend une étude d'héritabilité qui nécessitait :

- Un arbre généalogique : dans le but de suivre les caractères phénotypiques au cours des générations.
- Des caractères phénotypiques.

Des croisements des deux souches de parasites *Sm* BRE et *Sm* GH2 ont été effectués. Ces deux souches sont choisies pour les différences qu'elles présentent au niveau de certains caractères phénotypiques.

Les principales différences phénotypiques que présentent les deux souches :

- Phénotypes liés à des traits de vie :
 - (i) Les études de compatibilité effectuées au laboratoire montrent que les deux souches *Sm* BRE et *Sm* GH2 montrent des compatibilités (traduites par des prévalences et des intensités) qui sont différentes vis-à-vis des mollusques *Bg* BRE et *Bg* GUA (Theron *et al.* 2014).
- Phénotypes moléculaires, éventuellement associés aux phénotypes de trait de vie :
 - (ii) Les approches protéomiques comparatives effectuées entre *Sm* BRE et *Sm* GH2 ont permis de distinguer des candidats moléculaires qui sont différentiellement exprimés entre les deux souches de parasites. Parmi ces candidates, nous focalisons nos travaux sur les *SmPoMucs*, qui seraient également liés à la compatibilité entre le mollusque et le parasite (Roger *et al.* 2008a; Roger *et al.* 2008c).
 - (iii) Dans la partie 1 de ce chapitre, nous avons montré par une approche génétique comparative, notamment par séquençage des promoteurs de différents groupes de *SmPoMucs*, de très faibles variations nucléotidiques Single Nucleotide

Polymorphism (SNPs) entre les deux souches de parasite. Ces variations sont de l'ordre de 0 à 8 SNPs de différence sur des régions allant de 1 à 2 Kb (Perrin *et al.* 2013).

(iv) Egalement, nous avons montré que les modifications épigénétiques contrôlent la transcription des *SmPoMucs*. Cette étude a mis en évidence une différence de structure chromatinienne dans le promoteur de *SmPoMuc 3.1 (r1r2)*, qui montre également une différence de taux de transcription des ARNm entre *Sm BRE* et *Sm GH2* (Perrin *et al.* 2013).

Nous avons choisi *SmPoMuc 3.1 (r1r2)* comme candidat moléculaire pour cette étude pour son différentiel d'expression et de structure chromatinienne entre *Sm BRE* et *Sm GH2*. Ceci nous a permis, par une approche ciblée, d'étudier à la fois la transmission de la séquence de l'ADN (génétique), de sa structure chromatinienne (épigénétique) et de son phénotype moléculaire (transcription).

Il est important de noter que, outre ce marqueur *SmPoMuc*, un autre candidat présentant des différences de structure chromatinienne et d'expression entre *Sm BRE* et *Sm GH2* existe. Ce candidat est le gène *Smp_171100*, il a été identifié par une approche épigénétique globale comparative entre les deux souches de parasite (Lepesant *et al.* 2011).

Il code une protéine appartenant à une famille de métalloprotéases relativement conservée des invertébrés aux vertébrés (Bland *et al.* 2008). Chez *S. mansoni* et d'autres parasites nématodes, les métalloprotéases joueraient un rôle dans l'immunosuppression de leur hôtes. En effet, ils sont impliqués dans la dégradation des tissus de l'hôte par leur activité enzymatique zinc dépendante (Duvaux-Miret *et al.* 1992; Redmond *et al.* 1997).

Dans cette présente étude, la fonction de ce gène *Smp_171100* n'est pas particulièrement intéressante si ce n'est sa pertinence pour l'interaction avec l'hôte. Nous nous intéressons plutôt au différentiel d'expression et de structure chromatinienne que présente ce gène entre les deux différentes souches du parasite.

Pour cette raison, l'héritabilité génétique et épigénétique de Smp_171100 a été étudiée en parallèle de celle de *SmPoMuc 3.1* (r1r2).

II.2.2. Approches méthodologiques

1. Les modifications épigénétiques seraient-ils des sources des variants phénotypiques?
2. Quel est le poids relatif du « génotype » et de « l'épigénotype » dans l'expression du phénotype ?
3. Ces modifications épigénétiques suivent-ils une hérédité mendélienne?

II.2.2.1. Les croisements et génération des hybrides de *S. mansoni*

Les résultats de la première partie de ce chapitre ont montré qu'effectivement, les modifications épigénétiques pourraient être une source des variants moléculaires chez le parasite.

Afin de confirmer ce résultat et répondre aux deux questions suivantes, trois générations hybrides du parasite F1, F2 et F3 ont été réalisées. Il était essentiel d'aboutir la génération F3 dans le but de suivre l'hérédité de la chromatine.

En effet, selon les lois de Mendel, les miracidia de la génération F1 sont tous des hybrides hétérozygotes pour tous les loci. La génération F2 est une population de miracidia qui est hétérogène : pour 1 locus, nous avons 25% des miracidia génotype BRE, 25% des miracidia génotype GH2 et 50% miracidia génotype hétérozygote.

Les infestations mono-miracidiales dans notre modèle d'étude sont avantageuses. En effet, le parasite (miracidia) se développe dans le mollusque par des multiplications asexuées. Ceci permet d'obtenir une population clonale de milliers de cercaires ayant un seul sexe et un seul génotype.

En passant de la génération F2 à la génération F3, par des infestations mono-miracidiales, nous sélectionnons un miracidium de la population F2, et par conséquent un génotype particulier pour un locus donné. Ce génotype (les 2 allèles d'un locus) pourrait être *Sm BRE*, *Sm GH2* ou bien un hybride (avec les deux différents allèles). A ce stade de nos croisements, nous avons effectué le génotypage

des populations de cercaires, provenant chacune d'un seul miracidium F2. Nous avons pu sélectionner les parents F2 homozygotes pour notre gène candidat *SmPoMuc 3.1 (r1r2)*.

Le but de cette sélection c'était d'obtenir une ou plusieurs générations F3 ayant un génotype parental (*Sm BRE* et / ou *Sm GH2*). Ceci a permis d'étudier la transmission de l'épigénotype et du génotype à la fois.

1- Nous avons pu étudier la contribution de chacun (génotype et épigénotype) dans l'expression du phénotype moléculaire (transcription du gène).

2- Nous avons pu déterminer le type de l'héritabilité de l'épigénotype par rapport au génotype.

La figure 20 montre la procédure expérimentale des croisements des souches *Sm BRE* et *Sm GH2* et les générations obtenues au cours de la thèse. Au départ, des infestations mono-miracidiales ont été effectuées : *Sm BRE* sur son mollusque sympatrique *Bg BRE* et *Sm GH2* sur son mollusque sympatrique *Bg GUA*.

Cinq semaines après, le sexe de chacune des populations clonales de cercaires issues de chaque mollusque a été identifié. Après détermination du sexe, les parents ont été choisis, et cette étape a été suivie par les infestations des hamsters. Les croisements effectués sont les suivants :

Contrôle F0 : *Sm BRE* femelle x *Sm BRE* mâle

Sm GH2 femelle x *Sm GH2* mâle

Pour la F1 : *Sm GH2* femelle x *Sm BRE* mâle

Sm BRE femelle x *Sm GH2* mâle (un deuxième croisement a été utilisé comme réplicat biologique).

Chaque hamster a été infesté par 300 cercaires (200 cercaires mâles + 100 cercaires femelles). Six à dix hamsters ont été utilisés pour chaque croisement.

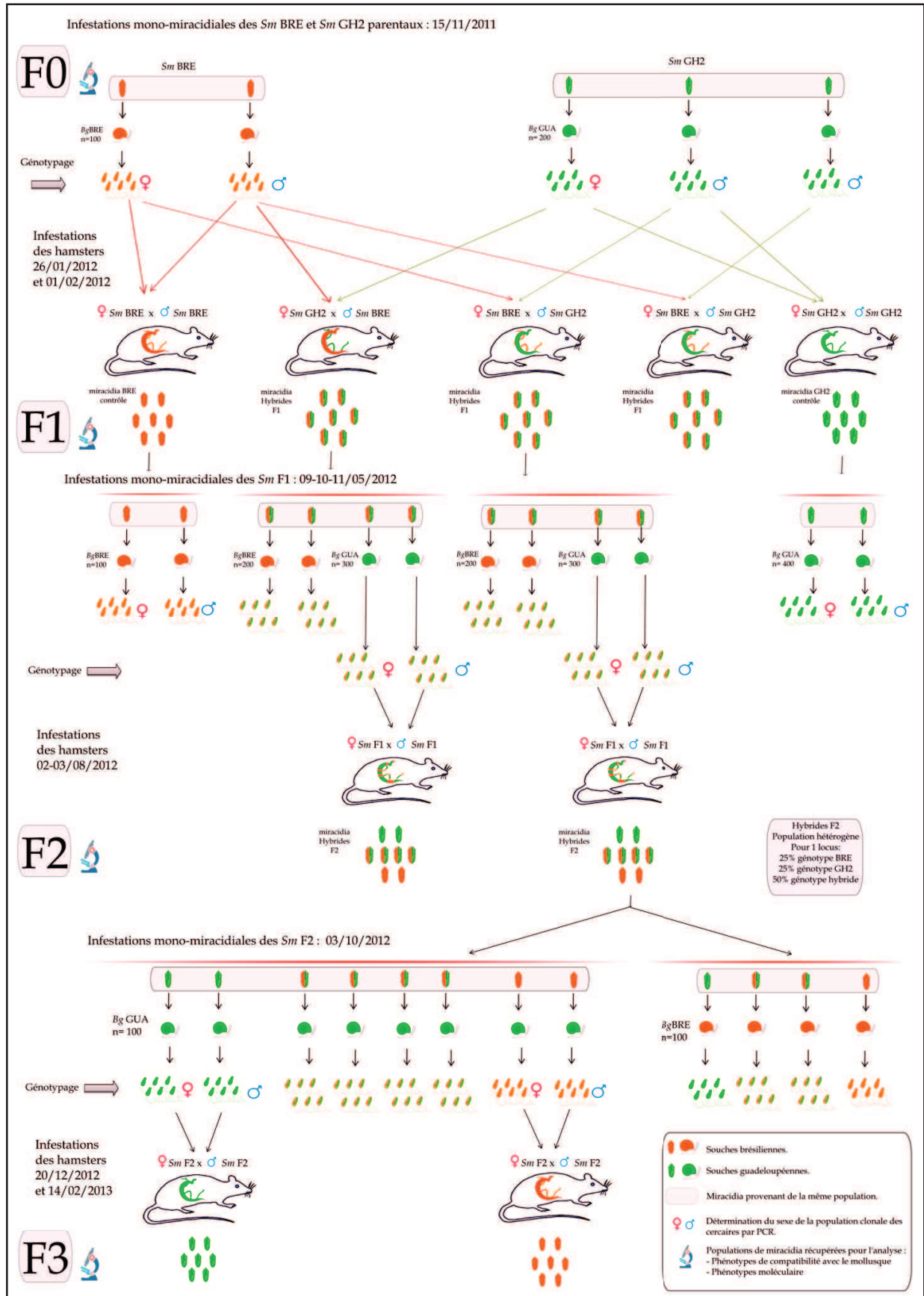


Figure 20 : Procédure expérimentale des croisements des deux souches du parasite *Sm* BRE et *Sm* GH2. Représentation schématique des expérimentations effectuées lors des croisements au cours de la thèse. La partie « Génotypage » indiquée dans ce schéma est explicitée dans les résultats non publiés plus loin et la Discussion de la thèse (chapitre IV).

Deux mois après les infestations, les foies des hamsters sont récupérés et broyés. Après l'éclosion des œufs, les miracidia sont récupérés et divisés en trois parties :

(i) Une partie a été utilisée pour étudier la compatibilité avec les mollusques. Pour cela, 27 à 50 mollusques de chacune des deux souches *Bg* BRE et *Bg* GUA ont été utilisés pour les miracidia issus de chaque croisement. Chaque mollusque a été exposé à 20 miracidia. Quinze jours après, les prévalences et les intensités ont été testées.

(ii) Une deuxième partie a été utilisée pour les analyses moléculaires. 1000 miracidia pour les analyses d'expression protéique par western blot, 1000 à 5000 miracidia pour les analyses de la transcription des ARNm, 10 000 miracidia pour l'étude de la structure chromatinienne par ChIP et 500 miracidia pour l'extraction d'ADNg et séquençage ciblé.

(iii) La troisième partie a servi pour les infestations mono-miracidiales sur les mollusques *Bg* BRE et *Bg* GUA dans le but d'aboutir la génération suivante.

La même démarche a été suivie jusqu'à la génération F3.

Il est essentiel de noter que les analyses moléculaires effectuées sur les miracidia de la génération F3 concerneraient uniquement la population verte de la figure 20 (Publication N°2).

II.2.2.2. Le traitement de *S. mansoni* à la Trichostatin A (TSA)

Nous avons déjà démontré que la TSA augmente la transcription de certains groupes de *SmPoMucs* (Cosseau *et al.* 2010; Perrin *et al.* 2013). Puisque les *SmPoMucs* sont supposés être des éléments importants pour la compatibilité avec le mollusque, nous avons voulu voir l'effet de la TSA sur le phénotype de compatibilité de *S. mansoni* avec le mollusque *B. glabrata*.

Pour cette étude, nous avons choisi un nouveau couple *Sm* LE / *Bg* BAR (des isolats géographiques différents mais du Brésil aussi), avec le couple *Sm* GH2 / *Bg* GUA.

En effet, les travaux d'André Théron ont montré que la souche *Sm* GH2 est la souche la moins compatible avec le mollusque parmi les différentes souches du parasite testées. En parallèle, la souche du mollusque *Bg* BAR est complètement résistante à la souche du parasite *Sm* GH2 (Théron *et al.* 2014). Ces observations nous ont menées à poser les deux questions suivantes :

- Est-ce que l'induction des variations épigénétiques par la TSA, va modifier la compatibilité de la souche du parasite *Sm* GH2 / *Bg* BAR ?
- Observons-nous des variations dans le profil d'expression protéique des *SmPoMucs* après le traitement ?

Pour cela, des foies d'hamsters (contenant les œufs), infestés par *Sm* GH2 et *Sm* LE ont été récupérés. Le traitement à la TSA a été effectué en trois fois (20 µM chaque traitement). Les deux premiers traitements ont été effectués sur les foies. La troisième dose a été ajoutée le lendemain sur les œufs, après le broyage des foies (figure 21 A).

L'étape suivante c'était l'étude des compatibilités. Pour cela, 20 à 23 mollusques souche *Bg* BAR ont été exposés à 20 miracidia chacun : *Sm* GH2 non traités, *Sm* GH2 traités, *Sm* LE non traités et *Sm* LE traités (figure 21 B).

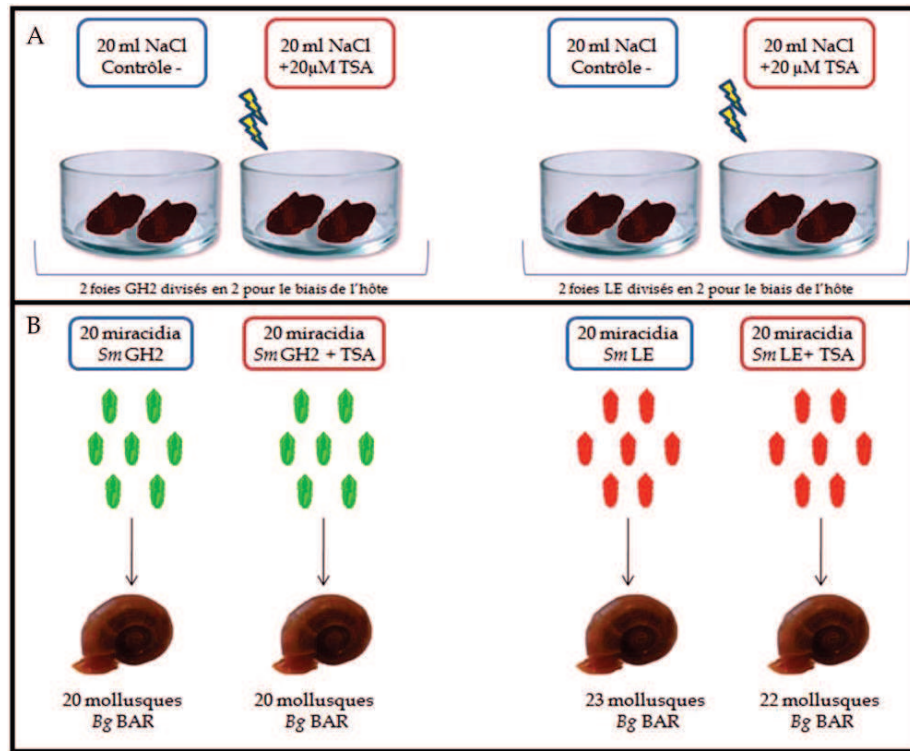


Figure 21 : Procédure expérimentale du traitement des *S. mansoni* à la Trichostatin A (TSA). (A) Traitement des foies des hamsters contenant les œufs de *Sm* GH2 et *Sm* LE à la TSA avec en parallèle, des foies non traités. (B) Infestation des mollusques *B. glabrata* souche BAR par les miracidia contrôles et traités : 20 miracidia / mollusque. 20 à 23 mollusques par condition. Les miracidia de *Sm* LE, isolat géographique différent, sont présentés en rouge.

II.2.3. Publication N°2

Epigenetic origin of adaptive phenotypic variants in the blood-fluke *Schistosoma mansoni*

Sara Fneich, André Theron, Céline Cosseau, Jérôme Buard, David Duval, Nathalie Arancibia, Jérôme Boissier and Christoph Grunau

Soumis à PLOS Biology (début novembre 2014).



Epigenetic origin of adaptive phenotypic variants in the blood-fluke *Schistosoma mansoni*

Authors and Affiliations

Sara Fneich 1)2)4), André Theron 1)2), Céline Cosseau 1)2), Jerome Buard 3), David Duval 1)2), Nathalie Arancibia 1)2), Jerome Boissier 1)2) and Christoph Grunau 1)2)

1) Université de Perpignan Via Domitia, Perpignan, F-66860, France

2) CNRS, UMR 5244, Ecologie et Evolution des Interactions (2EI), Perpignan, F-66860, France

3) CNRS, UPR1142, Institute de Génétique Humain (IGH), Montpellier, F-34396, France

4) Current address: INRA, UMR 1198 Biologie du Développement et Reproduction, Jouy en Josas, F-78352, France.

Abstract

Adaptive evolution is not possible without the generation of phenotypic variants. The origin of variation has been a central topic in evolutionary biology. It is now commonly accepted that standing genetic variation is the only cause of phenotypic variants. The term “mutant” is even often used interchangeable to designate a genetic or a phenotypic variation. However, epigenetic information is emerging as an alternative or complementary source of heritable phenotypic variation. It is currently not clear what the relative importance of genetics and epigenetics in generating these variants is. We used a host-parasite system to address this question. The human blood fluke *Schistosoma mansoni* can adapt relatively fast to new intermediate snail hosts. The interaction between parasite and mollusk is characterized by a compatibility polymorphism that makes the evolutionary dynamics accessible in the laboratory. The molecular marker for compatibility (infection success) is the expression pattern of a group of polymorphic mucins (*SmPoMuc*). We show here that chromatin structure changes at the *SmPoMuc* promotor regions are the principal cause for transcription polymorphism leading to phenotypic novelty and increase in infection success *i.e.* fitness. We establish here that epigenetic changes can be the major if not only cause of adaptive phenotypic variants in *Schistosoma mansoni*, suggesting that epimutations can provide material for adaptive evolution in the absence of genetic variation.

Author Summary

Evolution is a process that is based on three fundamental points: (1) the generation of phenotypic variants, (2) the selection of those variants that confer a reproduction advantage, and (3) the transmission of the capacity to reproduce these variants to the offspring. This latter point, the heritability, has so far been linked exclusively to the genes and random genetic mutations are thought to be the only source of heritable phenotypic variants. Since the middle of the last century we know that there exists an additional information that is closely linked to the genetic one and that influences gene activity: epigenetic information. Epigenetic information is influenced by the environment and can sometimes be transmitted to the offspring. It is therefore in theory possible that adaptation to new environments does not need genetic mutations to produce the phenotypic variants on which selection can act. While most scientists agree that epigenetic information exists, a sometimes heated scientific debate concerns the importance of it. Adaptation to new hosts allow parasites to spread. To control the spreading of the disease that is caused by the parasite one has to understand the mechanism of adaptation. This motivated our study of a parasitic flatworm, the agent of human biharzioses. For this parasite we show that epigenetic information is actually more important than genetic information to generate phenotypic variants of proteins that allow infecting its host. It might therefore be that epigenetics is an important support for rapid adaptation in general.

Blurb

In the human parasite *Schistosoma mansoni* epimutations are the cause for phenotypic variations that allow adaptation to new intermediate hosts.

Introduction

Adaptive evolution relies on the generation of heritable phenotypic variants on which selection can act. The origin of variation has puzzled researchers since Darwinian times [1]. There is now a relative broad consensus that genotype – environment interactions represent the major if not exclusive source of phenotypic novelty. However, this view has been challenged by theoretical considerations [2-4] [5] and it has been suggested that a substantial part of variability is the result of variations in the epigenetic component of the genome [6] [7]. It emerges a concept in which upon environmental changes, the epigenetic (“low-fidelity”) system allows a population to generate new phenotypes keeping genetic information invariant. This allows for exploring the fitness landscape after an environmental change and will “buy time” for the population. In case the new environment persists, genetically encoded phenotypic variants can emerge and the new phenotype can be genetically assimilated. Heritability of epiallels has been clearly established for numerous examples (e.g. [8] [9]) and several epigenetically encoded phenotypic characters were

described (e.g. [10] [11]). However, none of them is adaptive in the sense it provides a fitness gain under ecologically realistic conditions. We provide here, to the best of our knowledge, for the first time experimental evidence for such an epigenetically encoded adaptive phenotype. We chose a metazoan host/parasite system since selective pressures are strong in these systems, evolution is fast and effective population sizes are small.

Schistosoma mansoni is a human parasite causing the intestinal schistosomiasis. The life cycle is complex and requires the passage by two obligatory and successive hosts, the fresh water snail *Biomphalaria* spp. where asexual multiplication of the parasite occurs, and human or rodents as definitive hosts for sexual reproduction [12]. Specific strains of *S. mansoni* can only infect specific strains of *B. glabrata*, while other cannot be infected. This phenomenon is called compatibility polymorphism [13]. We had identified earlier molecular markers for compatibility by a comparative proteomic studies [14] [15]. These markers are the *Schistosoma mansoni* polymorphic mucins (*SmPoMucs*). *SmPoMucs* interact with the immune receptors of the invertebrate host [16]. They are encoded by a multigene family that is subject to different levels of regulation, including transcriptional, post-transcriptional and post-translational levels [15]. We recently demonstrated that transcriptional control is based on epigenetic mechanisms [17].

We used in the current study several inbred strains of *S. mansoni*. These strains show a specific heritable reaction norm in their compatibility with *B. glabrata* strains [18]. The compatibility character cannot be selected any further, *i.e.* offspring from those rare individuals that manage to infect a non-compatible mollusk shows the same reaction norm (host spectrum) as the original population (data not shown). Therefore, even though the strains are not clonal, they are pure lines. We show here that the first step to generate new phenotypic variants in a *S. mansoni* pure line is modification of the epigenetic information allowing for the generation of new phenotypes and rapid adaptation to new host strains.

Since we knew already that expression of key components for the host compatibility of *S. mansoni* is under epigenetic control we reasoned that the compatibility phenotype would be a suitable model to disentangle the effects of genetically and epigenetically heritable components. We applied a classical pedigree study and used two pure lines of *S. mansoni*: the Brazilian *SmBRE* strain and Guadeloupian *SmGH2* strain with their sympatric snail Brazilian snail *BgBRE* and Guadeloupian strain *BgGUA*, respectively. Both strains show strong differences in compatibility towards the same reference *B. glabrata* strain [18]. Compatibility phenotype corresponds to life history traits measured by prevalence and intensities. As mentioned above, *SmPoMucs* are key elements in the compatibility polymorphism between the miracidia of *S. mansoni* and its intermediate host *B. glabrata* [14]. The diversification

of this gene family is regulated by genetic [15] [19] [13] and epigenetic mechanisms [20] [17]. We therefore investigated expression of these genes.

Results

F3 Hybrids express more and different *SmPoMuc* variants than the parents

Thirty five to fifty snails from each strain were exposed individually to 20 miracidia. Prevalence was measured as the percentage of infected snails, and intensity corresponds to the number of mother sporocysts per snail. At each step miracidia were set aside to perform molecular biology analyses. The experiments were carried out with miracidia obtained from *SmBRE* x *SmBRE* and *SmGH2* x *SmGH2* crosses, F1 miracidia obtained from reciprocal crosses of *SmBRE* x *SmGH2*, F2 miracidia were obtained from F1 x F1 and F3 miracidia obtained from two different clonal population of F2 cercariae. The experimental scheme is summarised in figure 1. Quantitative RT-PCR shows that *SmPoMuc* 3.1(r1r2) is in average 18.3 fold more transcribed in *SmGH2* F0 than in *SmBRE* F0 strains (n=4) (figure 2A). This confirms our earlier results that major control of expression of these genes occurs at transcriptional level [17]. The hybrid lines F1 and F2 show higher transcription level compared to *SmBRE* F0 parent (data not shown). In the F3 generation, *SmPoMuc* 3.1(r1r2) is 11.3 fold higher expressed than in *SmBRE* F0 ($F_{3,18}=4.27$, $p=0.019$) while expression compared to *SmGH2* F0 is slightly, but statistically not significant, lower ($F_{3,18}=1.06$, $p=0.39$) (figure 2A). We wondered if this difference in transcription level concerns also other genes and tested 14 arbitrarily protein coding genes and transcribed repeats by qRT-PCR (supplementary table 1). In none of the cases we detected significant differences in transcription between the 4 generations (data not shown). We then used Western blots to test if *SmPoMuc* transcripts are translated. *SmPoMuc*s proteins were detected using an antibody that recognize the C-terminal conserved sequence by Western blot [16]. Proteins were extracted from 1,000 miracidia for each condition. We knew already that although there is very little nucleotide differences in the promotor regions of *SmPoMuc*s family genes [17], the expression profiles between the parental strains *SmBRE* F0 and *SmGH2* F0 were different [16], and we confirmed this here by Western blot (figure 3A). F1 and F2 hybrid miracidia express a combined profile that contains all bands of F0 parents. The F3 generation has an expression profiles similar to that of the F1 and F2 hybrids, but also new *SmPoMuc*s variants were expressed (figure 3A). If *SmPoMuc* expression would follow Mendelian segregation we would expect in F3 a profile similar to the F1 and F2 hybrid or to one of the parents. Since *SmPoMuc*s are key markers for host-parasite compatibility we expected the hybrids with a different *SmPoMuc* profile to show changes in compatibility. We therefore investigated their capacity to infest different snail strains.

F3 Hybrids show higher fitness than the parents

Prevalences in *BgBRE* are 96-100% for *SmBRE* F0 and 4-6% for *SmGH2* F0. Prevalences increase significantly (Fisher exact test, $p < 0.0001$) compared to *SmGH2* in F1, F2 and F3 to reach maximal values in F2 and F3 that are statistically not different from *SmBRE* F0 in all hybrid generations. Intensity, *i.e.* the mean number of mother sporocysts per snail *BgBRE* is 7.1 ± 0.25 for *SmBRE* ($n=34$) and 1 ± 0 for *SmGH2* ($n=50$). Similar to prevalence, intensity values increase significantly compared to *SmGH2* in F1, F2 and F3 (figure 4A). In *BgGUA* snails prevalences are 80-83% and 55-60% for parental miracidia *SmBRE* F0 and *SmGH2* F0, respectively. As for *BgBRE*, infectivity rates increase in F1, F2 and F3 hybrids to achieve maximal values statistically not different from *SmBRE* F0 ($p < 0.05$, Fisher exact test) for all hybrid generations. However, they are higher than the *SmGH2* F0 parent (figure 4A). For intensities on *BgGUA*, there is no significant difference between *SmBRE* and *SmGH2* parents, but we observe an increase of intensity values across generations (figure 4A). Consequently, all crosses of the two different strains of the parasite *S. mansoni* show increasing infection success of hybrids over three generations F1, F2 and F3, *i.e.* an increase in global fitness. Interestingly, while the *SmPoMuc* expression (molecular) phenotype is additive, compatibility phenotypes in F2 and F3 hybrids show uniparental dominance, and for both *SmPoMuc* patterns and infection success we see non-mendelian segregation. Consequently, we explored other mechanisms responsible for fitness increase. Since we had earlier shown that chromatin structure controls *SmPoMuc* expression [17], we decided to study segregation of chromatin marks at the *SmPoMuc* locus that shows strongest chromatin structure and expression differences between *SmBRE* and *SmGH2*: *SmPoMuc* 3.1 (r1r2).

Chromatin structure in *SmPoMuc*3.1 shows non-mendelian segregation

SmPoMuc promoter sequences do not show any methylated cytosines [17]. Therefore, we focus our analysis on histone modifications in the promoter regions by Chromatin ImmunoPrecipitation (ChIP) using two different antibodies that recognize histone H3 acetylated on lysine 9 (H3K9ac) and histone H3 tri-methylated on lysine 9 (H3K9me3). ChIP was followed by qPCR analysis to quantify the immunoprecipitated DNA. Since acetylation and methylation at H3K9 are mutually exclusive, we express chromatin status here as ratio of H3K9me3/H3K9ac. *SmPoMuc* 3.1(r1r2) promoter shows a very different chromatin structure between the two F0 parents *SmBRE* and *SmGH2* (figure 2B). This is correlated with differential expression of this gene between the two parents. In F3, chromatin structure is intermediate between F0 parents (figure 2B). The promoter regions of *SmPoMuc* 3.1 (r1r2) contain 3 SNPs between *SmBRE* F0 and *SmGH2* F0. To identify the genotypes at *SmPoMuc* 3.1 (r1r2) we sequenced DNA of miracidia from all generations *SmBRE* F0, *SmGH2* F0, F1, F2 and F3. As expected, F1 and F2 hybrid populations are heterozygous with two different parental alleles. The F3 clone was homozygous for

the *SmGH2* genotype. Consequently, the *SmPoMuc* 3.1 (r1r2) *SmGH2* genotype had acquired an epigenotype that is statistically different from *SmGH2* and approaches those of F0 *SmBRE* while not being identical. One might argue that this is part of a genome-wide change in chromatin structure. We therefore investigated another locus for which we had earlier identified differences in H3K9ac enrichment between *SmBRE* and *SmGH2* [21]: Smp_171100.

Smp_171100 codes for a putative M13 metallo-endopeptidases located on *Schisto_mansoni*.Chr_2 22,414,795-22,470,759. The function of the gene is not of particular importance here. The gene is not anymore listed in the latest version of the genome annotation but we have confirmed presence of an transcript by RT-PCR. We had shown earlier that chromatin differences exist [21] in the gene body between *SmBRE* and *SmGH2* and we used them here as epigenetic markers for both strains. We confirmed our findings that *SmBRE* F0 and *SmGH2* F0 show different H3K9me3/H3K9ac ratios. In F3, the ratio is identical to *SmBRE* F0 (figure 5B). These differences in chromatin structure are correlated with transcription differences (figure 5A). We sequenced a region of the Smp_171100 gene sequence containing 6 SNPs between *SmBRE* and *SmGH2*. As expected and as in *SmPoMuc* 3.1 (r1r2), F1 and F2 miracidial populations are heterozygous. The F3 population (the same population that exhibits the *SmGH2* alleles for *SmPoMuc* 3.1 (r1r2)) is homozygous *SmBRE* in Smp_171100.

In conclusion, while in the *SmPoMuc* 3.1(r1r2) locus, the *SmGH2* genotype had changed epigenotype in F3 (figure 5A), in Smp_171100 we see co-segregation of epigenotype and genotype (both are *SmBRE* in F3). In other words, for the two loci in which to our knowledge exist a clear epigenetic difference between *SmBRE* F0 and *SmGH2* F0 we distinguish two modes of chromatin heritability, one case of non-mendelian inheritance (*SmPoMuc*), and one case of mendelian inheritance and cosegregation of genotype and epigenotype (Smp_171100).

Pharmacological induction of chromatin structure changes leads to transcription of new *SmPoMuc* variants and renders incompatible (avirulent) strain compatible (virulent)

All lines of evidence indicated that it was the H3K9me3/ac ratio that controlled specifically *SmPoMuc* expression and that the change in chromatin structure resulted in new phenotypic variants based on new combinations of *SmPoMuc*. This in turn led probably to the increase in infection success of the hybrids. If this was true, then perturbation of H3K9me3/ac ratio by other means should also lead to new *SmPoMuc* variants and should have an impact on parasite-host compatibility. Locus-specific epigenetic engineering of *SmPoMuc* is not yet feasible. We therefore opted for pharmacological treatment and treated eggs of *S. mansoni* with the histone deacetylase inhibitor Trichostatin A (TSA). We had shown earlier that treatment with

20 μ M TSA increased the number alternative transcripts (number of polymorphic bands detected by RT-PCR) in *SmPoMuc* but does not alter transcription of control genes [20]. Also in other systems it was shown that TSA modulate transcription of around 5% of genes, and that genes can be up but also down regulated by TSA [22] [23] [24] [25]. In other words, TSA does not induce a global increase in transcription but will activate or inactivate certain loci only. We treated the *S. mansoni* eggs three times before hatching with 20 μ M TSA. This dose has negligible cytotoxic effect [20]. For this study we used again two pure lines of *S. mansoni*: the Brazilian *SmLE* strain and *SmGH2* strain with the Brazilian snail *BgBAR* (sympatric with *SmLE*). Western blot (figure 2B) shows that TSA treatment leads to expression of new *SmPoMuc* protein variants in the miracidia, *i.e.* the already known transcription variation translates into changes also on the protein level. We subsequently investigated the compatibility phenotype using prevalences and intensities without and with TSA treatment. Twenty to twenty three snails from *BgBAR* strain were exposed individually to 20 miracidia. Without treatment, prevalences in *BgBAR* are 95.65% (n=23) for *SmLE* and 5% for *SmGH2* (n=20). Prevalence increase significantly (Fisher exact test, $p < 0.05$) for *SmGH2* after treatment with TSA reaching 35% (n=20), and an increase to 100% was also observed in *SmLE* (n=22). Furthermore, the intensity values increase for *SmGH2* from 2 ± 0 to 3.29 ± 0.5 after TSA treatment, and for *SmLE* from 5.91 ± 0.73 to 9.55 ± 0.87 (figure 4B).

In conclusion, treatment with an epimutagenic agent does not only lead to changes in the *SmPoMuc* expression patterns but also allows an nearly incompatible parasite to become compatible with a new snail host.

Discussion

In parasite – host interactions, infection success can be used as an estimation for parasite fitness. It can be measured as number of infected host individuals (prevalence) and number of successful infection events per host individual (intensity). Using these parameters, we show here that after a hybridisation event between two different *S. mansoni* strains, the offspring shows increased fitness. Increasing of fitness of hybrids is not surprising [26] [27], but in general such hybrid vigor or heterosis is attributed to increased genetic diversity in the offspring, leading to new phenotypes with higher fitness. Here we demonstrate that this is not the only mechanism by which phenotypic variability increases in hybrids. At least for the *SmPoMuc* loci, epigenetics plays the most important role. This makes sense in light of the fact that nucleotide diversity between the two parental strains at these loci is extremely low [17]. We did not address the question by which mechanism chromatin structure is modified. It could be that the *SmBRE* locus is paramutagenic for *SmGH2*; it could also be that the *SmBRE* epigenotype shows some sort of dominance effect.

Further work will be needed to elucidate this. Using pharmacological induction of epimutations we also observed new *SmPoMuc* variants. Taken together, we thus confirmed our earlier findings that epigenetic mechanisms control transcription of the *SmPoMuc* genes and we show now that epigenetic information is indeed the origin of phenotypic novelty in these loci. The exact function of *SmPoMucs* is still elusive and for the moment we did not establish a clear causal relationship between new *SmPoMuc* variants and infection success. Nevertheless, our earlier studies had firmly established *SmPoMuc* variability as a key marker for compatibility between *S. mansoni* and its intermediate host [15] [16]. We conclude that strain hybridization and TSA treatment lead to changes in the epigenetic information that establish novel developmental trajectories leading to new and more phenotypic variants in the population. This allows the parasite population to explore the fitness landscape. If new matching phenotypes between parasites and mollusk host are produced, the latter can be infected, *i.e.* the phenotype is adaptive. Our finding fit perfectly into theoretical models that predict such a function for the epigenetic information [4] [2], but are to the best of our knowledge the first experimental evidence for an epigenetic basis of adaptive evolution.

Materials and Methods

Ethics statement

Our laboratory has permission A 66040 from both French Ministère de l'agriculture et de la pêche and French Ministère de l'Education Nationale de la Recherche et de la Technologie for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87-848 and 2012201-0008) for the experimenters. Housing, breeding and animal care followed the national ethical requirements.

Culture of *Schistosoma mansoni* parents and hybrids strains

We used in this study two different strains of *S. mansoni*: *SmBRE* and *SmGH2*. Each strain was maintained in its sympatric mollusc *Biomphalaria glabrata* strain as intermediate host and in hamsters (*Mesocricetus auratus*) as definitive host as described previously [28]. Three generations of hybrids of *SmBRE* and *SmGH2* were produced as follows. Monomiracidial infections were performed within each strain of mollusk *BgBRE* and *BgGUA*. We obtained clonal populations of cercariae after 4 weeks. Sex was determined by PCR as described before [29] [30]. Strain hybrids were produced by infection of hamsters with 300 cercariae: 200 males from a clonal cercarial population and 100 females from another clonal population for each generation. Different combinations of parental cercariae were used for each generation in order to generate biological replicates. Three months later, eggs were collected from hamster livers and hatched in spring water. Miracidia were divided

into three parts: (i) one part was used for life trait studies; (ii) a second part was used for pedigree (iii) a third part was concentrated by sedimentation on ice for 30 minutes and stored for molecular analysis.

Sex identification of cercariae

Four weeks after the monomiracidial infestation of mollusk for each generation BRE F0, GH2 F0, F1 and F2, four clonal cercariae were selected from each mollusk in order to determine the sex by PCR as described previously [29] [30]. After DNA extraction from single cercariae, PCR was performed using two pairs of primers, one control primers that amplifies Rhodopsin (Smp_scaff001984 (49840-50016)) on male and female, and one female specific primers that amplifies SmWSPP2 (Smp_scaff002739 (2682-3046)) only on female (supplementary table 1). Knowing sex of cercarial population by mollusk, we chose the two parents male and female in order to infest hamsters and produce the next generation.

Pharmacological induction by Trichostatin A treatment

For this treatment we used two different strains of *S. mansoni*: *SmLE* and *SmGH2*. After dissection of 4 infested hamsters (2 for each strain), livers were collected and divided into two parts to eliminate the host bias. Two half livers for each strain *SmLE* and *SmGH2* were incubated in 20 mL NaCl as control. For treatment, two half livers for each strain *SmLE* and *SmGH2* were incubated in 20 mL NaCl adding two times 20 µl of Trichostatin A (TSA) (InvivoGen met-tsa-5) dissolved in ethanol to 20 mM. All livers were incubated overnight at 4°C. The next day, directly after grinding livers, a third treatment was performed on eggs from treated livers for *SmLE* and *SmGH2*. Miracidia (non treated *SmGH2*, treated *SmGH2*, non treated *SmLE* and treated *SmLE*) were divided on two parts (i) for prevalence and intensity analysis and (ii) for Western blot analysis).

Compatibility of *S. mansoni* with mollusk hosts

Single mollusks (n = 30 to 50) were exposed to 20 miracidia each in 5 ml of spring water overnight at 24-25°C. Prevalence was measured 15 days post-exposure, by determining the rate of infected mollusks over the entire mollusks that were exposed to parasites. The intensities were evaluated by counting the number of mother sporocysts that developed within infected mollusk as previously described [31].

Western blot

Total proteins were prepared using 1,000 miracidia from each condition (parental strains and hybrid strains) by incubating in 30 µl Lämmli buffer, 5 minutes at 99°C. Fifteen µl were separated by 10% SDS-PAGE gel electrophoresis before being blotted on a nitrocellulose membrane (Trans-Blot turbo, Bio-Rad). The membrane was blocked with 5% skimmed dry milk in TBST (TBS buffer containing 0.05% Tween 20) by stirring one hour at room temperature and then incubated with the primary

antibody “anti-*SmPoMuc*” [16] diluted 1/500 in TBST for 90 minutes at room temperature. The membrane was then incubated with secondary antibody (peroxidase conjugated, purified anti-rabbit IgG) diluted 1/5000 in TBST for 1 hour and washed three times with TBST. Finally, the blot was subjected to the ChemiDoc MP Imaging system- BioRad) using the ECL reagents.

For Trichostatin A treatment, we prepared total proteins from 200 miracidia of non treated GH2, 600 miracidia from treated GH2, 500 miracidia from non treated LE and 500 miracidia from treated LE. For western blot, we followed the same procedure with *SmPoMuc*s antibody as for the hydrids. To calibrate, the membrane was dehybridized 15 minutes with Re-Blot plus strong solution (Millipore) and re-incubated with the primary antibody “anti-actin” (Thermo scientific, MA 1-744) as control diluted 1/200 in TTBS for 2 hours at room temperature. Then, it was incubated with secondary antibody (peroxidase conjugated, purified anti-mouse IgG+M, life technologies) diluted 1/500 in TBST for 1 hour at room temperature.

RNA extraction, reverse transcription and quantitative PCR

Messenger RNAs were isolated from 500 to 5,000 miracidia from each strain using the Dynabeads® mRNA DIRECT Micro Kit (Invitrogen). The samples were put directly into 100 µl lysis buffer at -80°C, and processed according to the manufacturer’s instructions. After washing, the samples were resuspended directly in 20 µl of DNase treatment mix (Ambion RNA by Life Technologies DNA-free). cDNA were synthesized from 13 µl of the total mRNA preparation, in a final volume of 20 µl using the RevertAid Premium First Strand cDNA Synthesis Kit by Thermo Scientific. Quantitative PCR analyses were carried out using 2.5 µl of cDNA diluted 1/10 in a final volume of 10 µl (1.5 µl H₂O, 0.5 µM of each primer, 5 µl of master mix), using a LightCycler® 480 Real-Time Instrument, and 2.5 µl of mRNA diluted 1/20 as negative control for the specific exon junctions amplifications. Alpha-tubuline (α -tub) was used for normalizing as internal control. Primer sequences are listed in supplementary table 1. The following protocol was used: denaturation, 95°C 10 min, amplification and quantification (45 cycles), 95°C for 10s, 60°C for 5s, 72°C for 20s; melting curve, 65-97°C with a heating rate of 0.11 C/s and continuous fluorescence measurement, and cooling step to 40°C. For each reaction, the crossing point cycle threshold (Ct) was determined using the “second derivate” method of the LightCycler® 480 Software release 1.5.0. Reactions were carried out in duplicate then the mean Ct was calculated.

The amplification of a unique band of each locus was verified by size separation on a LabChip GX a capillary electrophoretic system.

Chromatin Immunoprecipitation (ChIP) followed by qPCR

Native chromatin immunoprecipitation was performed as described before [32]. The following antibodies against histone isoforms were used to precipitate chromatin in miracidia: Abcam anti-H3K9me3 (ab8898, Lot 733951) and Millipore anti-H3K9ac (07-352, Lot DAM1576933). Immunoprecipitated DNA was extracted by phenol/chloroform protocol and analyzed by quantitative PCR using specific primers for *SmPoMuc 3.1(r1-r2)* promoter (primer sequences in supplementary table 1) and for the *Smp_171100*. The amount of target DNA recovered in the immunoprecipitated fraction was quantified by calculating the percent of input recovery (% IR) normalized with the percent input recovery obtained with α -tubulin gene. The percent input recovery of the bound immunoprecipitated fraction for each amplicon was calculated by the following formula: $\% \text{ input recovery} = 100 \times E^{(\text{Ct}(\text{input}) - \text{Ct}(\text{IP}_{\text{Bound}}))}$. The percent background was calculated by the following formula: $\% \text{ background} = 100 \times E^{(\text{Ct}(\text{input}) - \text{Ct}(\text{C}_{\text{Bound}}))}$, where E is the primer efficiency designed to amplify the amplicon, $\text{Ct}(\text{IP}_{\text{Bound}})$ is the Ct of the bound fraction obtained in the immunoprecipitated sample, $\text{Ct}(\text{C}_{\text{Bound}})$ is the Ct of the bound fraction obtained in the negative control (fraction without antibody), and $\text{Ct}(\text{input})$ is the Ct of the unbound fraction obtained in the negative control. It represents the quantity of chromatin that was used for the study minus the fraction that bound non-specifically to the protein A Sepharose beads. Finally, the ratios of the H3K9me3 on H3K3Ac for each gene was calculated using formulas $(2^{\text{ct}}(\alpha\text{-tub})/2^{\text{ct}}(\text{SmPoMuc 3.1(r1r2)}))$ and $2^{\text{ct}}(\alpha\text{-tub})/2^{\text{ct}}(\text{Smp_171100})$.

DNA extraction, PCR amplification and specific sequencing around SNPs

Genomic DNA for *SmBRE F0*, *SmGH2 F0*, *F1*, *F2* and *F3* miracidial population, was prepared by the incubation 3 hours at 55°C with 500 μl of lysis buffer (20 mM Tris/Cl Ph 8; 1 Mm EDTA; 100 mM NaCl; 0.5% SDS) and 0.15 mg of proteinase K. The samples were extracted twice with equal volumes of phenol/chloroform, followed by two extractions of equal volumes of chloroform. DNA was precipitated with equal volumes of isopropanol/NaOAc (3M, pH 5.2) at room temperature. After centrifugation and washing with 1 ml of 70% ethanol, the pellet was dissolved in 50 μl of 1 mM Tris/Cl pH 8.

In order to determine if the chromatin structure follows the mendelian transmission or not (co-segregate with the allele or not), we focus specifically on sequences where we determined the chromatin structure by ChIP-qPCR for *SmBRE F0*, *GH2 F0*, *F1*, *F2* and *F3* miracidial population. We identified 3 SNPs between *SmBRE F0* and *GH2 F0* for *SmPoMuc 3.1(r1r2)* promoter sequence and 6 SNPs for *Smp_171100* sequence using Sequencher software. Primers including SNPs for the two loci were designed and the regions were amplified by qPCR. For the *SmPoMuc 3.1(r1r2)* promoter, qPCR was applied on a 9kb PCR product containing the 2kb region upstream of the

transcription start site (TSS) [17], and for Smp_171100 sequence, qPCR was applied directly on genomic DNA for *SmBRE* F0, *SmGH2* F0, F1, F2 and F3 miracidial population. qPCR products were sequenced using Eurogentec facilities (<http://www.eurogentec.com/life-science.html>). We checked and aligned nucleotide sequence manually using Sequencher and BioEdit softwares.

Statistical analysis

Fisher exact test was used to compare prevalences and T-test to compare intensities. For the statistical differences in the amount of transcripts, we used ANOVA with Dunnett post-hoc test using the parental strain as control group.

Acknowledgments

We would like to thank Guillaume Mitta, Benjamin Gourbal, and Richard Galinier for enlightening discussions. We thank also Anne Rognon and Bernard Dejean for providing valuable technical support. Special thanks to G. Oliveira (Instituto Nacional de Ciência e Tecnologia em Doenças Tropicais, Fundacao Oswaldo Cruz, Brazil) for the generous gift BgBAR / SmLE strains. This work received funding from the ANR 2010 BLAN 1720 01 (EPIGEVOL).

References

1. Darwin C (1868) The variation of animals and plants under domestication: John Murray.
2. Klironomos FD, Berg J, Collins S (2013) How epigenetic mutations can affect genetic evolution: model and mechanism. *Bioessays* 35: 571-578.
3. Maynard Smith J (1990) Models of a dual inheritance system. *Journal of theoretical biology*: Elsevier. pp. 41-53.
4. Pal C, Miklos I (1999) Epigenetic inheritance, genetic assimilation and speciation. *J theor Biol*: Elsevier. pp. 19-37.
5. Waddington CH (1942) The epigenotype. *International journal of epidemiology* 41: 10-13.
6. Tal O, Kisdi E, Jablonka E (2010) Epigenetic Contribution to Covariance Between Relatives. *Genetics* 184: 1037-1050.
7. Frésard L, Morisson M, Brun J-M, Collin A, Pain B, et al. (2013) Epigenetics and phenotypic variability: some interesting insights from birds. *Genet Sel Evol* 45: 16.
8. Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nature genetics* 23: 314-318.
9. Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, et al. (2014) Mapping the epigenetic basis of complex traits. *science* 343: 1145-1148.
10. Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, et al. (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* 441: 469-474.

11. Manning K, Tör M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature genetics* 38: 948-952.
12. Morgan JAT, Dejong RJ, Adeoye GO, Ansa E, Barbosa CS, et al. (2005) Origin and diversification of the human parasite *Schistosoma mansoni*. *Molecular Ecology: Wiley Online Library*. pp. 3889-3902.
13. Mitta G, Adema CM, Gourbal B, Loker ES, Theron A (2012) Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Developmental and Comparative Immunology* 37: 1-8.
14. Roger E, Mitta G, Mone Y, Bouchut A, Rognon A, et al. (2008) Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach. *Molecular and Biochemical Parasitology* 157: 205-216.
15. Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, et al. (2008) Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). *PLoS neglected tropical diseases: Public Library of Science*. pp. e330.
16. Moné Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, et al. (2010) A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS One: Public Library of Science*. pp. e813.
17. Perrin C, Lepesant JM, Roger E, Duval D, Fneich S, et al. (2013) *Schistosoma mansoni* Mucin Gene (SmPoMuc) Expression: Epigenetic Control to Shape Adaptation to a New Host. *PLOS Pathogens: Public Library of Science*. pp. e1003571.
18. Theron A, Rognon A, Gourbal B, Mitta G (2014) Multi-parasite host susceptibility and multi-host parasite infectivity: A new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni* compatibility polymorphism. *Infection, Genetics and Evolution* 26: 80-88.
19. Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, et al. (2008) Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* 157: 217-227.
20. Cosseau C, Azzi A, Rognon A, Boissier J, Gourbiere S, et al. (2010) Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* - a possible kick-off for adaptive host/parasite evolution. *Oikos* 119: 669-678.
21. Lepesant JM, Grunau C, Cosseau C (2011) Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. *Memorias do Instituto Oswaldo Cruz* pp. 823-830.
22. Lopez-Atalaya JP, Ito S, Valor LM, Benito E, Barco A (2013) Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. *Nucleic acids research: gkt590*.
23. Halsall J, Gupta V, O'Neill LP, Turner BM, Nightingale KP (2012) Genes are often sheltered from the global histone hyperacetylation induced by HDAC inhibitors. *PloS one* 7: e33453.

24. Andrews KT, Gupta AP, Tran TN, Fairlie DP, Gobert GN, et al. (2012) Comparative gene expression profiling of *P. falciparum* malaria parasites exposed to three different histone deacetylase inhibitors. *PloS one* 7: e31847.
25. Ehrenkaufer GM, Eichinger DJ, Singh U (2007) Trichostatin A effects on gene expression in the protozoan parasite *Entamoeba histolytica*. *BMC genomics* 8: 216.
26. Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. *Trends in plant science: Elsevier*. pp. 427-432.
27. Baranwal VK, Mikkilineni V, Zehr UB, Tyagi AK, Kapoor S (2012) Heterosis: emerging ideas about hybrid vigour. *Journal of Experimental Botany: Soc Experiment Biol*. pp. 6309-6314.
28. Théron A, Pages J-R, Rognon A (1997) *Schistosoma mansoni*: Distribution Patterns of Miracidia among *Biomphalaria glabrata* Snail as Related to Host Susceptibility and Sporocyst Regulatory Processes. *Experimental Parasitology: Elsevier*. pp. 1-9.
29. Julien Portela CG, Céline Cosseau, Sophie Beltran, Christelle Dantec, Hugues Parrinello and, Boissier J (2010) Whole-genome in-silico subtractive hybridization (WISH) - using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics* 11: 387.
30. Julie M.J. Lepesant JB, Déborah Climent, Céline Cosseau, Christoph Grunau (2013) Female biased sex-ratio in *Schistosoma mansoni* after exposure to an allopatric intermediate host strain of *Biomphalaria glabrata*. *Experimental Parasitology* 135: 350-356.
31. Théron A, Gerard C (1994) Development of accessory sexual organs in *Biomphalaria glabrata* (Planorbidae) in relation to timing of infection by *Schistosoma mansoni*: consequences for energy utilization patterns by the parasite. *Journal of molluscan studies: Malacological Soc London*. pp. 25-31.
32. Cosseau C, Azzi A, Smith K, Freitag M, Mitta G, et al. (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Molecular and Biochemical Parasitology* 166: 70-76.

Financial Disclosure

This section should describe sources of funding that have supported the work. Please include relevant grant numbers and the URL of any funder's Web site. Please also include this sentence: "The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." If this statement is not correct, you must describe the role of any sponsors or funders, and amend the aforementioned sentence as needed.

Competing Interest

The authors declare not to have competing financial interests.

Abbreviations

SmPoMuc: *Schistosoma mansoni* polymorphic mucins

SmBRE: *Schistosoma mansoni* strain Brazil BRE

SmLE: *Schistosoma mansoni* strain Brazil LE

SmGH2: *Schistosoma mansoni* strain Guadeloupe GH2

BgBRE: *Biomphalaria glabrata* strain Brazil BRE

BgGUA: *Biomphalaria glabrata* strain Guadeloupe GUA

BgBAR: *Biomphalaria glabrata* strain Brazil BAR

ChIP: Chromatin Immunoprecipitation

H3K9ac: histone H3 acetylated on lysine 9

H3K9me3: histone H3 tri-methylated on lysine 9

TSA: Trichostatin A

Figure legends

Figure 1: Crossing scheme for pedigree study. After monomiracidial infection of their sympatric hosts in F0, crosses within each generation were produced in each generation until F3.

Figure 2: Boxplot of transcription level and chromatin structure in the promoter of *SmPoMuc3.1* (r1r2). Two technical replicates for each biological replicate (n). (A) Relative transcription of *SmPoMuc3.1* (r1r2) compared to alpha-tubuline measured by RT-qPCR. (B) H3K9me3 to H3K9ac ratio at the promoter of *SmPoMuc3.1* (r1r2). Transcription level and chromatin structure do not co-segregate.

Figure 3: Typical examples of expression profiles of *SmPoMucs*. Western blot experiments were performed using a standardized method on proteins extracts from 1,000 miracidia. (A) *SmPoMucs* expression in different generations. The genotypes at the *SmPoMuc 3.1* (r1r2) locus are indicated above. (B) *SmPoMucs* expression in *SmLE* with and without TSA treatment. After TSA treatment, new variants are clearly detectable (indicated by *).

Figure 4: Prevalence and intensity of *S. mansoni* infection in *B. glabrata* snail strains. (A) Prevalence is expressed as % of infected snails, intensity as mean number of SpI sporocysts per snail. (B) Prevalence and intensity in snail strain *BgBAR* for two parasite strains *SmGH2* and *SmLE*, with and without TSA treatment.

Figure 5: Boxplot of transcription level and chromatin structure in the gene body of Smp_171100. Two technical replicates for each biological replicate (n). (A) Relative transcription compared to alpha-tubuline measured by RT-qPCR. (B) H3K9me3 to H3K9ac ratio. Transcription level and chromatin structure co-segregate.

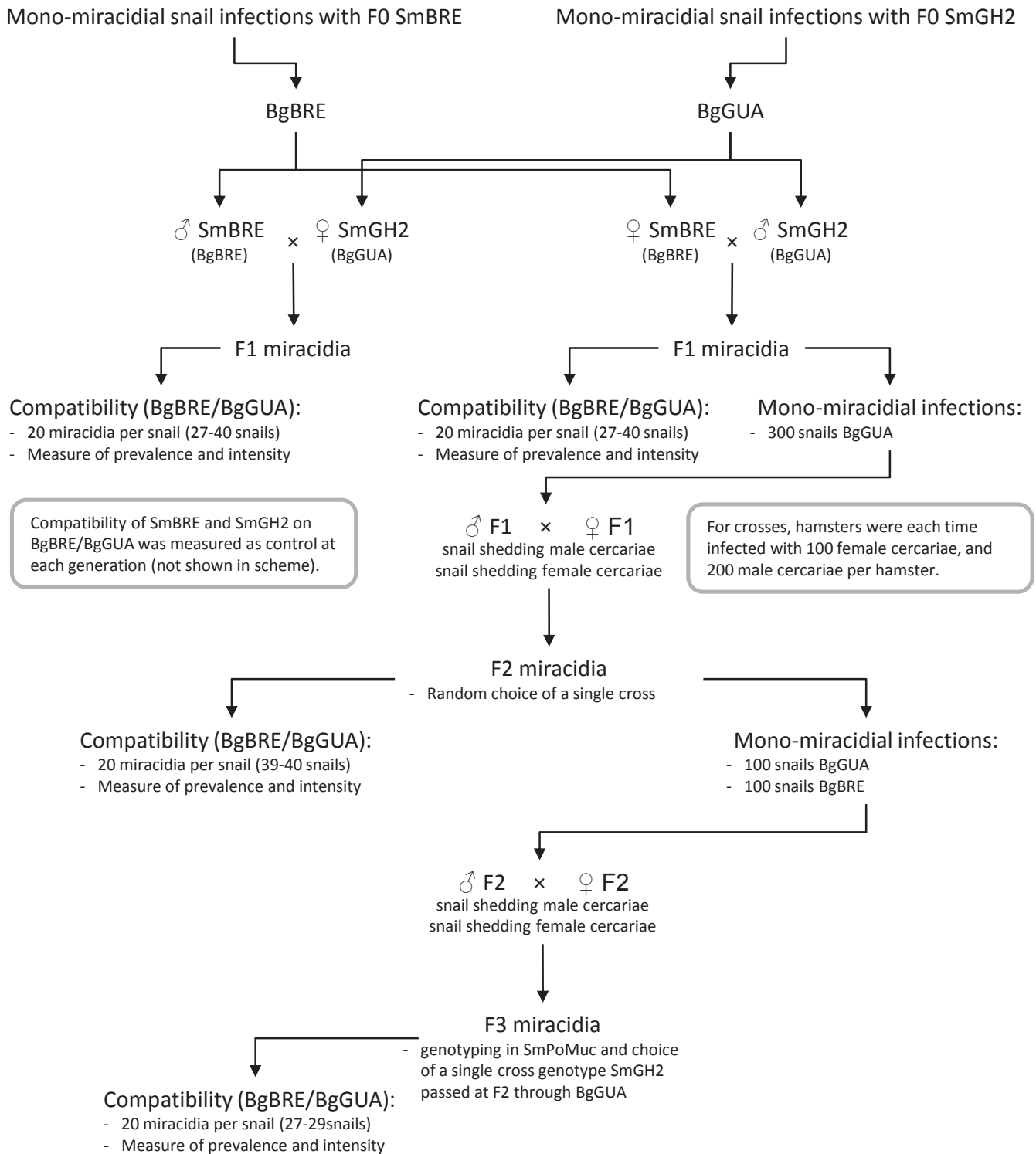
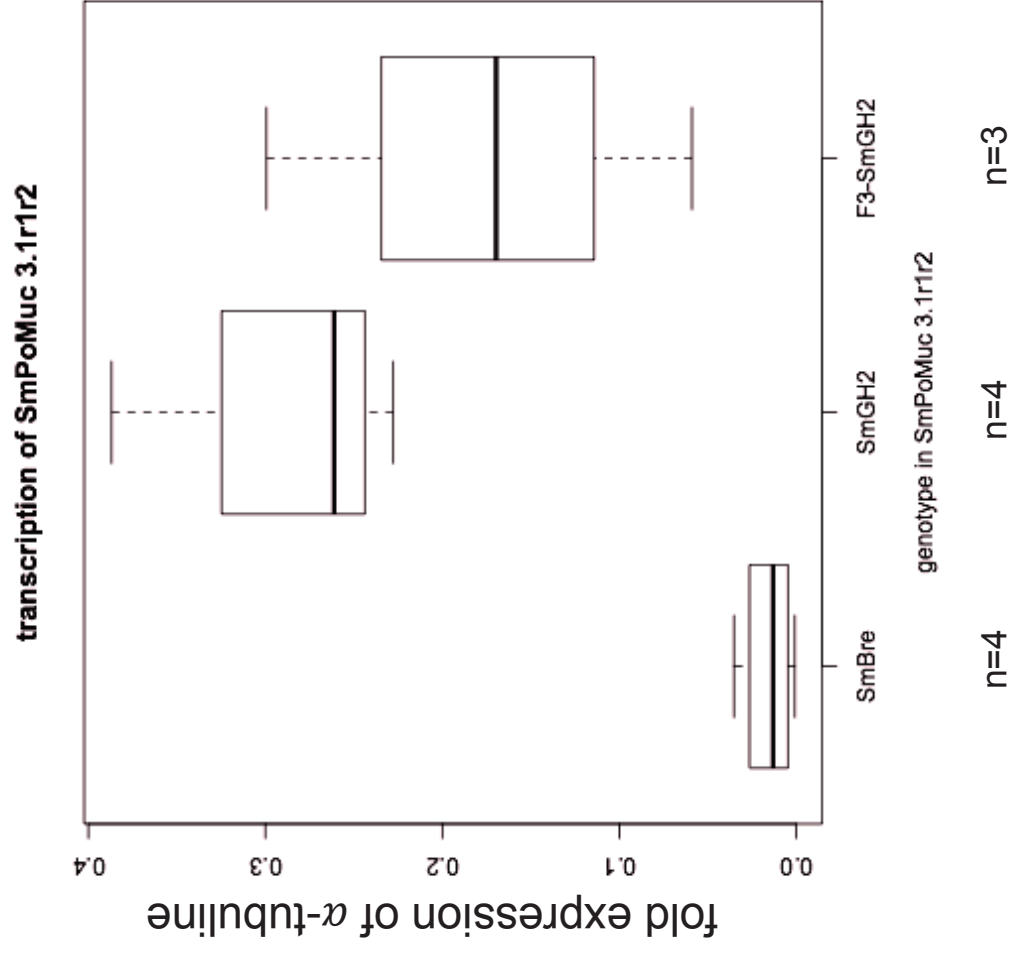


Figure 1: Crossing scheme for pedigree study. After monomiracidial infection of their sympatric hosts in F0, crosses within each generation were produced in each generation until F3.

A



B

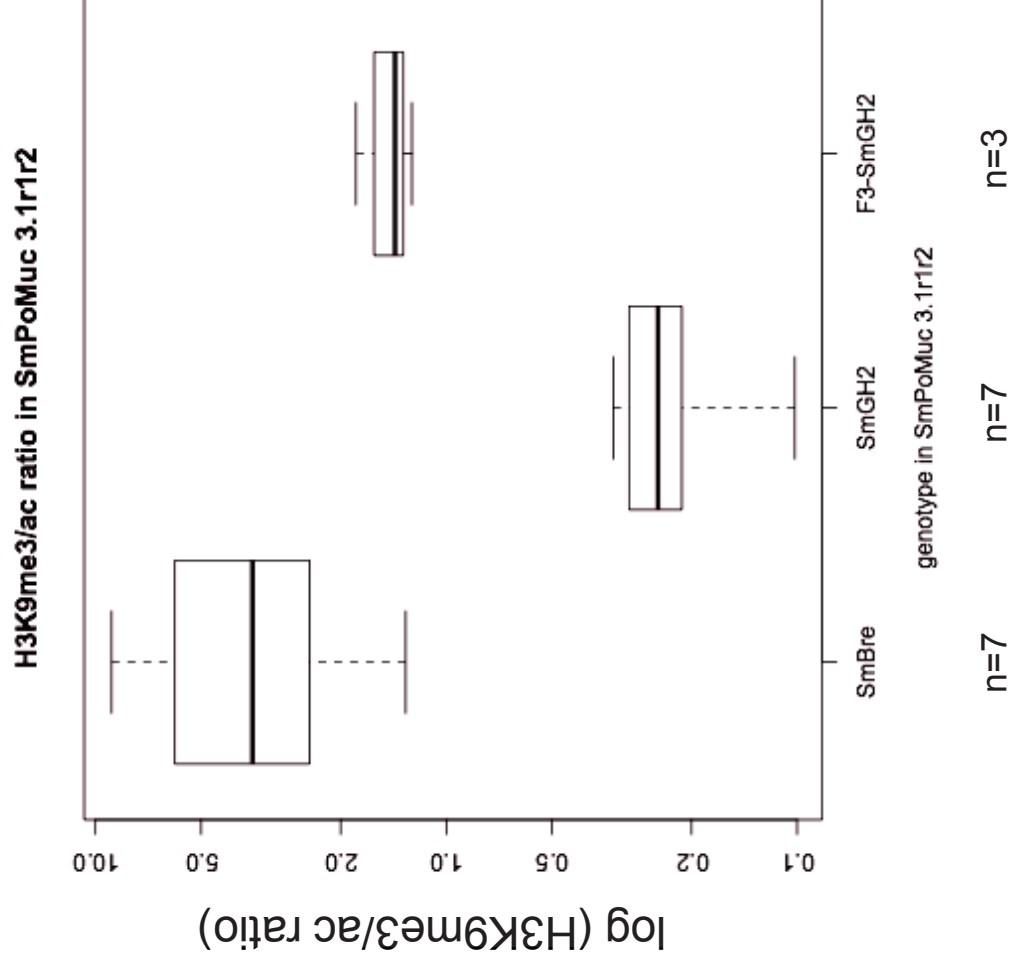


Figure 2: Boxplot of transcription level and chromatin structure in the promoter of *SmPoMuc3.1 (r1r2)*. Two technical replicates for each biological replicate (n).
 (A) Relative transcription of *SmPoMuc3.1 (r1r2)* compared to alpha-tubuline measured by RT-qPCR. (B) H3K9me3 to H3K9ac ratio at the promoter of *SmPoMuc3.1 (r1r2)*. Transcription level and chromatin structure do not co-segregate.

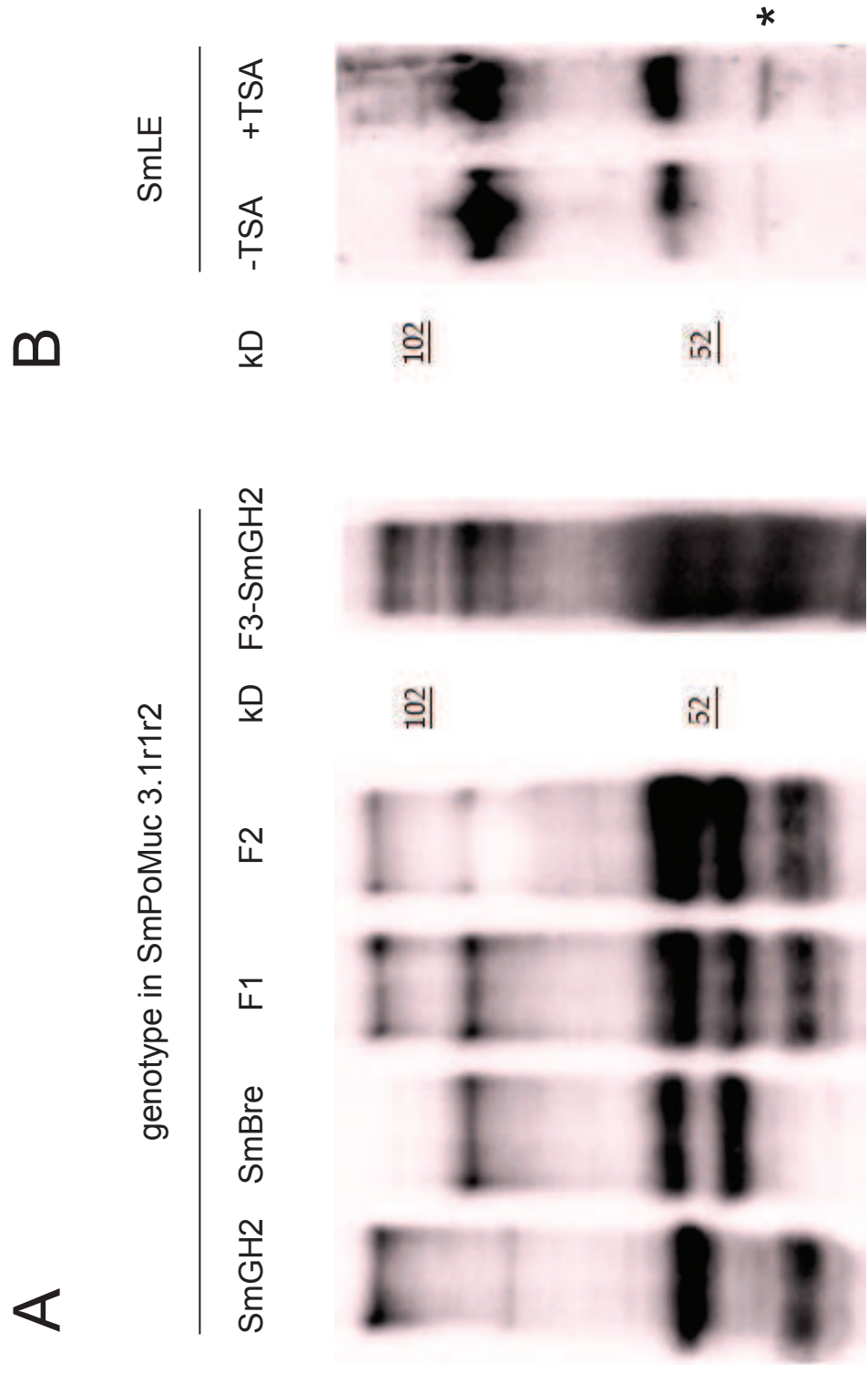
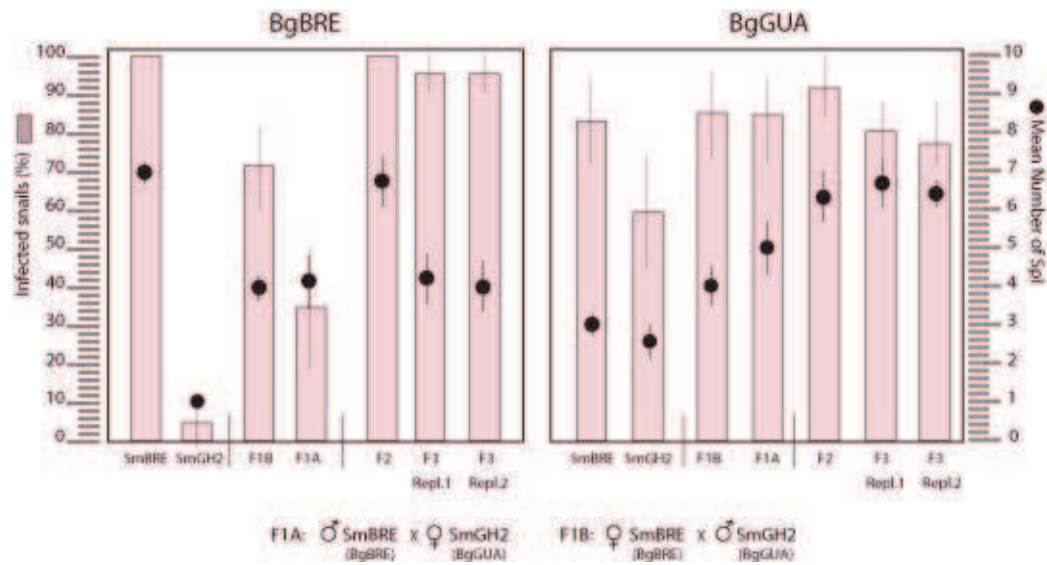


Figure 3: Typical examples of expression profiles of *SmPoMucs*. Western blot experiments were performed using a standardized method on protein extracts from 1,000 miracidia. (A) *SmPoMucs* expression in different generations. The genotypes at the *SmPoMuc* 3.1 (r1r2) locus are indicated above. (B) *SmPoMucs* expression in *SmLE* with and without TSA treatment. After TSA treatment, new variants are clearly detectable (indicated by *).

A



B

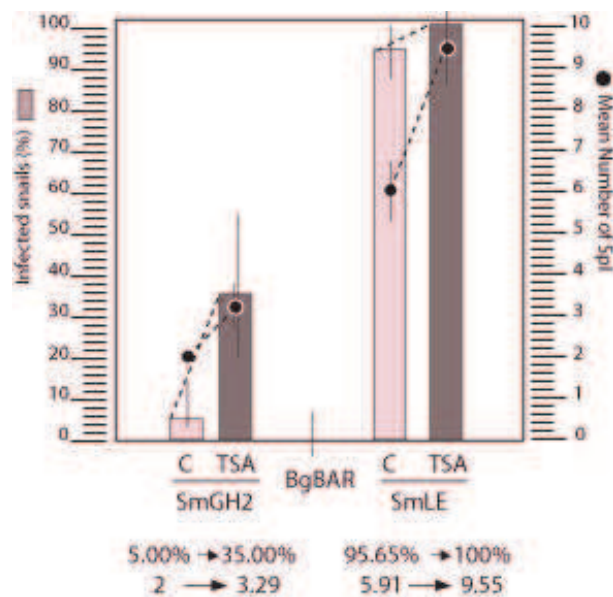
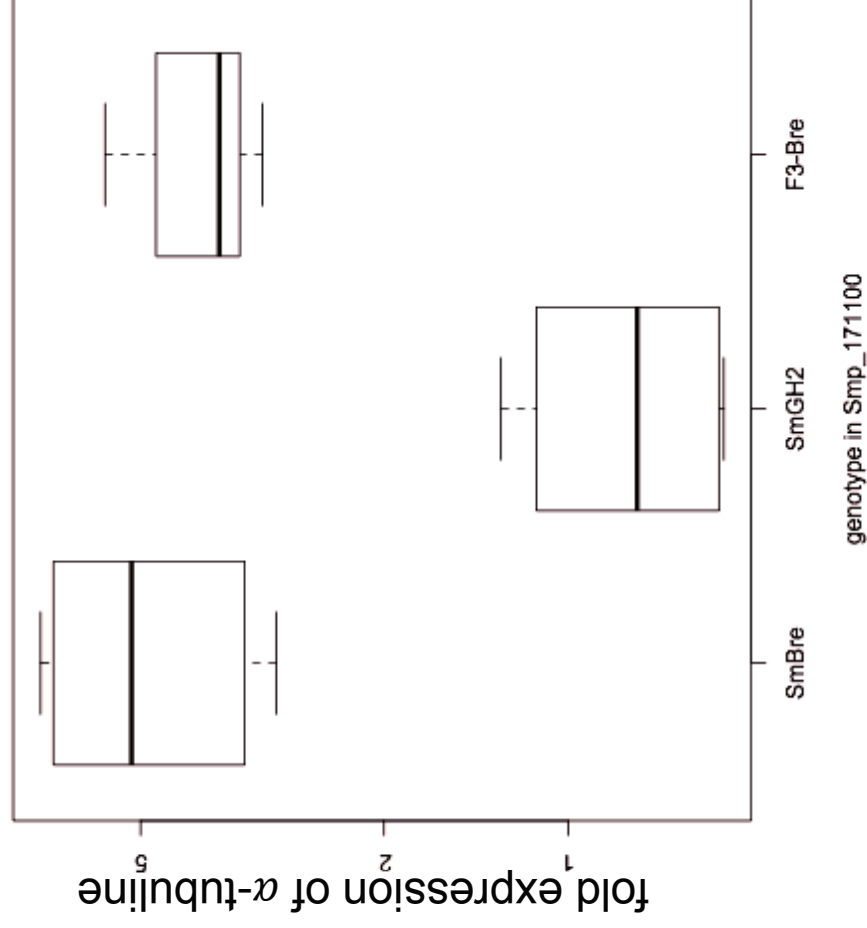


Figure 4: Prevalence and intensity of *S. mansoni* infection in *B. glabrata* snail strains. (A) Prevalence is expressed as % of infected snails, intensity as mean number of Spt sporocysts per snail. (B) Prevalence and intensity in snail strain BgBAR for two parasite strains SmGH2 and SmLE, with and without TSA treatment.

A

transcription of Smp_171100



B

H3K9me3/ac ratio in Smp_171100

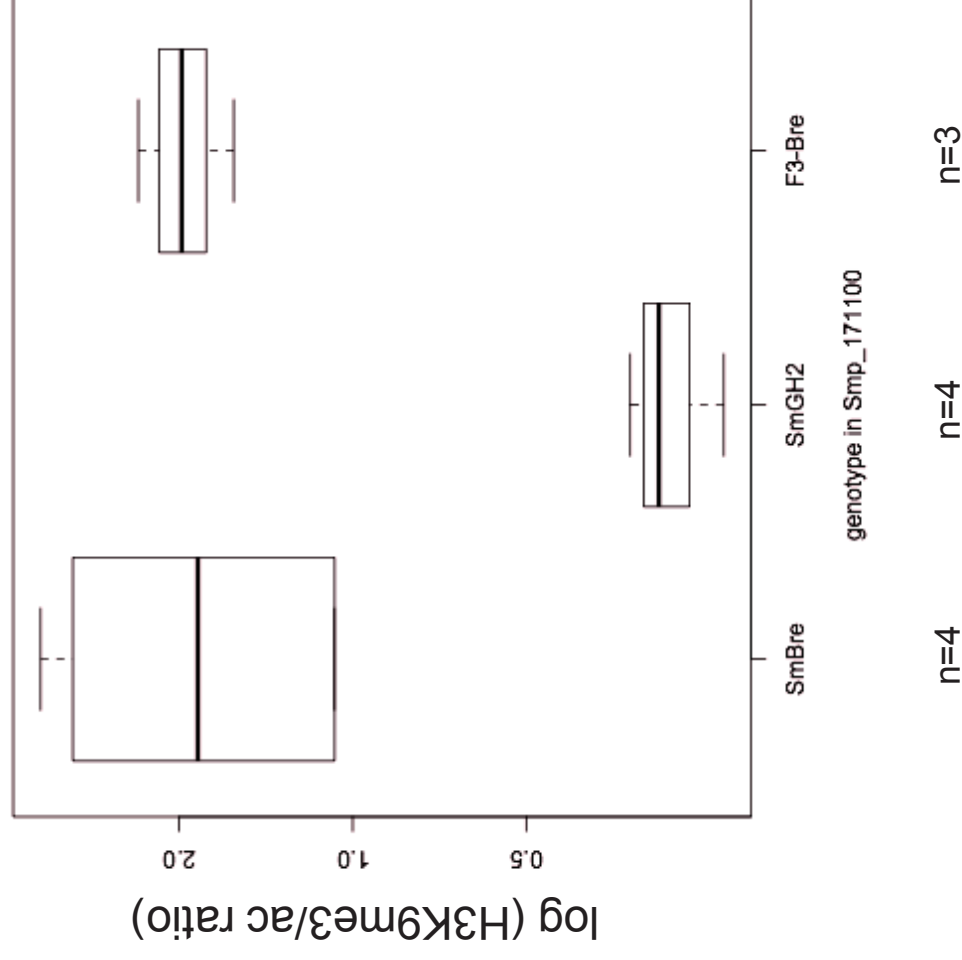


Figure 5: Boxplot of transcription level and chromatin structure in the gene body of Smp_171100. Two technical replicates for each biological replicate (n). (A) Relative transcription compared to α -tubuline measured by RT-qPCR. (B) H3K9me3 to H3K9ac ratio. Transcription level and chromatin structure co-segregate.

Supplementary table 1: Primers used in this study

Primer used for sex identification of cercariae

Target	Name	Forward (5'-3')	Reverse (5'-3')	Amplicon length (bp)
Rhodopsin	Smp_scaff001984.F Smp_scaff001984.R	GACGGCCACACTAAAG	AGTAAAATGGTCACTGCTAT	191
SmWSPP2	Smp_scaff002739.F Smp_scaff002739.R	TGTTTCGAATTTACACTT	CATTCACAGTTTGGCGAACA	391

Primer used for expression analysis by qPCR

Target	Name	Forward (5'-3')	Reverse (5'-3')	Amplicon length (bp)
α-tubuline	aTub2.F aTub2.R	TAGCGCACCATCGAAGCGC A	GCTTTCATGGTTGACAACGAGGCC AT	65
<i>SmPoMuc</i> 3.1(r1r2)	URG/B.F <i>SmPoMuc</i> 3 E2/E3.R	CGAACCAACAGGTGACCTC GC	CTGGAAGTGTGGTTCGCTC	113
MSPP tot	MSPPtot. F MSPPtot. R	GAATTGGAATCGAAGGCG	TGTTTCATCTTTGGGTAGTGT	182
<i>SmPoMuc</i> c1	smpomuc 1.E8/9.F smpomuc 1.E12.R	CGAAAGTGCTTACATCGCTG	CGAATAGGCTTCTTTATGTTG	190
<i>SmPoMuc</i> c2	smpomuc 2.E4.F smpomuc 2.E8.R	CCAGATCCACAACAGCACC	ACTTTCGACGTCATCAAGAAG	240
GST	GST.B.F GST.B.R	CAAAGGCCTTGTAACAACCA ACTCGT	CGTTTCGCCACACATCGCCG	97
GAPDH	GAPDH.F GAPDH. R	TGGCCGTGGAGCGATGCAA A	CGTCTGGTGTTGGGACGCGG	122
R564 W5	R=564.F R=564.R	ACCGCGTCTCCCCTCTTCAG T	AGGGAGAGCGCACACCATGA	137

R879 W4	R=879.F R=879.R	CGGTCATGGAGGCGTGGTCA	CTCGTCGTACCATGACAATCAGCA	120
smPDI	SmPDI.F SmPDI.R	GGGATTTATCAAGGATACGG ACTC	CACCAAGGAGCATACAGTTTGAC	
smProt1	smProt1.F smProt1.R	TGGCTATCGTAAATAAAGA AGGAA	TCAGCAATAGACCAATCACATTT	72
cox 2	Cox2.F Cox2.R	AGATGGATGCGGTTCCAGGT CG	AGCATGACCCACACCGCATAAC	104
Smp 152710.1	Smp_1527 10.1A.F Smp_1527 10.1A.R	ACAGCTCTAGTTGTCGACCA AACATC	ACGCGGTCAACGTAAGGGCA	132
Smp 054160	Smp_0541 60A.F Smp_0541 60A.R	CGGACGCGGACGTGCTGAA T	CAGGCAATCGTCCGCCTGGA	114
Smp 158110.1	Smp_1581 10.1A.F Smp_1581 10.1A.R	GCAAAGCTGGTGGTTTGGGG C	TGCATGACCTTCCTGTTCGTC AAG	90
Spo/AF1 09180	Spo4.F Spo4.R	TGTGTCGTAGGTGCTATG	TGCCTAAATCTTCCTTCCG	204
Ef1a	Ef1a2.F Ef1a2.R	AAACTGAAGGCTGAACGCG AACG	AGCCGCAACAATCAGCATTGCAC	158

Primer used for ChIP analysis by qPCR

Target	Name	Forward (5'-3')	Reverse (5'-3')	Amplicon length (bp)
α-tubuline	ATub.F ATub.R	AGCAGTTAAGCGTTGC AGAAATCA	TGACGAGGGTCACATTTACCA T	65
<i>SmPoMuc</i> 3.1(r1r2)	<i>SmPoMuc</i> 3.1(r1r2)	ATGATTTTACGAGAGGT TCAGT	GGTGCACTATTCAGCCTAGTCA	101
Metallopeptidase	Smp171100.F Smp171100.R	CCTGGAATGGAGAGCT CGTATGACA	GGAGTGTGGATTGAACGCGTCG	115

Primers used for PCR amplification and specific sequencing around SNPs

Target	Name	Forward (5'-3')	Reverse (5'-3')	Amplicon length (bp)

SmPoMuc 3.1(r1r2)	exon2.F exon1.R	TTCTTAGCACTACCCA AAGATGAAC	AGAGAATAATTTTCTTGTTTCATT CTTC	9000
	R1R2-SNP1.F R1R2-SNP1.R	CACGTGTTGCAAAATA ACCG	GAAATCCCGGAAGCACTGTA	204
	R1R2-SNP2.F R1R2-SNP2.R	CCGCAGACGAGTCCA ATAGT	AGCAACCGATGATTTGCATT	115
	R1R2-SNP3.F R1R2-SNP3.R	TTTCCAAAACGTTGGA CCAT	CGATTGTAAATTTGGAACGAT	170
	R1R2-SNP4-5.F R1R2-SNP4-5.R	CATGAACAGGGAAAT CACGA	AAAACCATGTTGTTCCGAGG	127
Metallope ptidase Smp_17110 0	Smp171100.F Smp171100.R	CCTGGAATGGAGAGCT CGTATGACA	GGAGTGTGGATTGAACGCGTCG	115

II.2.4. Résultats complémentaires non publiés

La virulence des hybrides au niveau de l'hôte définitif :

Le croisement des deux souches du parasite *Sm* BRE et *Sm* GH2 a engendré des parasites hybrides virulents. Leur virulence a été observée aussi bien sur l'hôte intermédiaire *B. glabrata* que sur l'hôte définitif *Mesocricetus auratus*.

Au niveau de l'hôte intermédiaire, la virulence des parasites a été évaluée par l'augmentation de la compatibilité du parasite vis-à-vis du mollusque. Les traits de vie qui ont été étudiés incluent les prévalences et les intensités (Publication N°2).

Les hybrides de *S. mansoni* ont été également virulents vis-à-vis de leur hôte définitif (figure 22). Aucun cas de mortalité n'a été observé chez les hamsters infestés par les miracidia de la génération F1. Ces hamsters F1, comme les hamsters infestés par les souches parentales (*Sm* BRE F0 et *Sm* GH2 F0), ont résisté pendant plusieurs mois. De même, le nombre moyen de miracidia par hamster F1 ne montrait pas de différences significatives par rapport à celui des souches parentales.

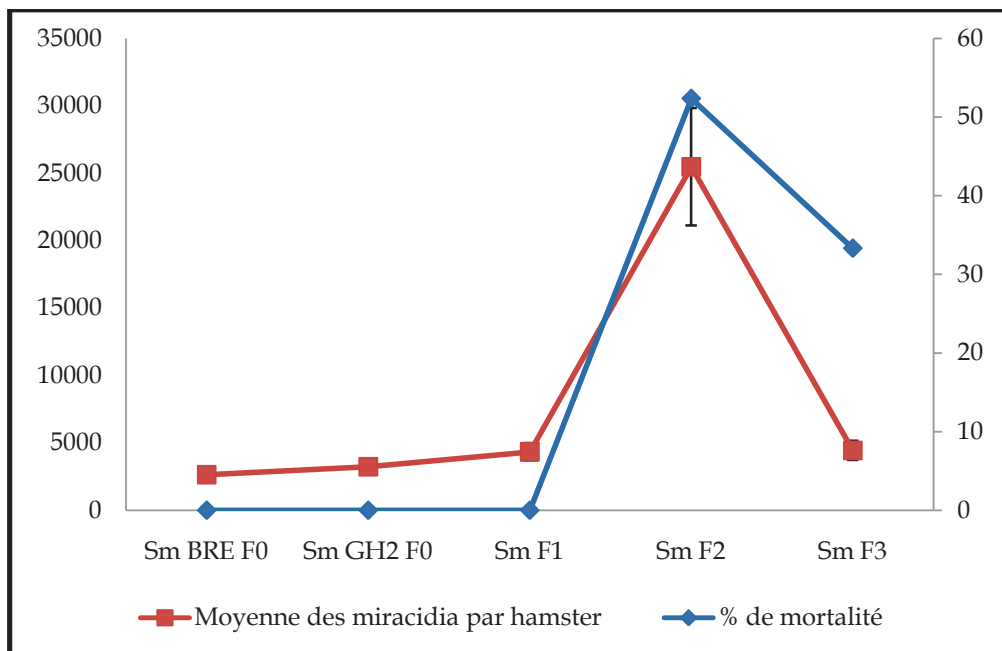


Figure 22 : La virulence des hybrides de *S. mansoni*. A gauche, est présenté le nombre moyen des miracidia obtenu à partir d'un hamster. A droite, c'est le pourcentage des hamsters morts 7 à 8 semaines après infestation par *S. mansoni*.

Cependant, la génération des *Sm* F2 était la génération hybride la plus virulente, qui a provoqué la mortalité de la moitié des hamsters infestés. La mortalité a été observée 7 à 8 semaines après infestation. Ainsi, une inflammation aigüe a été repérée au niveau des intestins de tous les hamsters F2. En parallèle de la mortalité, le nombre moyen de miracidia par hamster était très élevé par rapport à celui des souches parentales et des autres générations hybrides. Le nombre maximal de miracidia a atteint les 60000 par hamster F2. Il est important de rappeler que tous les hamsters de toutes les générations ont été infestés de la même façon, et avec le même nombre de cercaires.

Ensuite, la virulence a baissé de nouveau à la génération F3. Le taux de mortalité ainsi que le nombre moyen de miracidia par hamster ont diminué par rapport à ceux de la génération F2.

Par conclusion, le phénomène de « vigueur hybride » ou « hétérosis » a été observé après le croisement de deux souches de parasite provenant de deux origines géographiques différentes. Les hybrides ont montré une performance supérieure au moins par rapport à l'une des souches parentales (*Sm* GH2).

En plus de ce résultat, nous remarquons aussi que le phénotype « virulence », se transmet d'une manière non mendélienne au cours des générations. En effet, selon les lois de Mendel, la génération F2 est une population hétérogène constituée par (25% génotype maternel, 25% génotype paternel et 50% génotype hybride). Par conséquent, avec ces pourcentages, nous attendons de voir des phénotypes F2 plus proches de ceux des parents que les phénotypes F1 ou bien des phénotypes similaires à ceux de la F1 au minimum. Cependant, ce n'était pas le cas. Nous déduisons que les caractères phénotypiques transmis aux générations montrent ainsi une exception aux lois de Mendel.

Génotypage des générations hybrides :

Afin de s'assurer des génotypes de différentes générations de *S. mansoni*, les souches parentales *Sm* BRE et *Sm* GH2 et les souches hybrides *Sm* F1, *Sm* F2 ont été génotypés.

Pour les génotypages, nous avons utilisés des marqueurs génétiques spécifiques à chacune des souches *Sm* BRE et *Sm* GH2. Neuf marqueurs microsatellites neutres variables entre les deux souches (a, c, d, e, f, g, i, k, et m) et trois marqueurs génétiques variables (o, p et s) ont été utilisés (tableau 2) (Beltran *et al.* 2008; Lepesant *et al.* 2011).

L'ADNg a été extrait à partir de différentes populations clonales de cercaires par la technique de phénol/Chloroforme (décrite sur le site <http://methdb.univ-perp.fr/epievo/>).

Les populations de cercaires qui ont été génotypées, sont les populations parentales (population mâle et population femelle) qui ont servi pour chacun des croisements (voir la figure 20).

La figure 23 présente les génotypes des 10 populations de cercaires différentes. Les deux populations *Sm* BRE mâle et *Sm* BRE femelle ont servi pour le croisement qui a engendré les miracidia souche BRE. Les deux populations *Sm* GH2 mâle et *Sm* GH2 femelle ont servi pour donner les miracidia souche GH2. Egalement *Sm* F1 mâle et *Sm* F1 femelle ont servi pour donner les miracidia de la génération F2. Et finalement, *Sm* F2 mâle et *Sm* F2 femelle (ceux qui sont représentées ici) ont été utilisés pour le croisement qui a engendré la population de miracidia de la génération F3. C'est les miracidia F3 qui présentent le génotype GH2 pour *Sm*PoMuc 3.1 (r1r2) (Publication N°2).

	Marqueurs génétiques	Parasite <i>Sm</i> BRE	Parasite <i>Sm</i> GH2	Localisation	Position
a	R95529 (R95529)	273	240	Chromosome 4	430416 - 430746
b	SmC1 (AF325694)	290	290	Chromosome 2	24279889 - 24280250
c	SmD57 (AF202967)	296	274	Chromosome 2	6218862 - 6219172
d	SmBR16 (L04480)	337	341	Chromosome 4	12801807 - 12802217
e	SmDO11 (AF325698)	331	375	Chromosome 6	13534547 - 13534902
f	SmDA28 (AF325695)	115	91	Chromosome 2	2667958 - 2668285
g	SmBR1 (L81235)	154	156	Chromosome W	55209712 - 55209924
h	SmS7-1 (AF330105)	184	184	Chromosome 3	5798564 - 5798586
i	SmD28 (AF202966)	241	244	Chromosome 3	19511435 - 19511969
j	SCMSMOXII (M85305)	295	295	Chromosome W	46345753 - 46346176
k	SmBR10 (DQ448293)	133	109	Chromosome 2	2667949 - 2668282
l	L46951 (L46951)	168	168	Chromosome 6	3701533 - 3701674
m	SmBR13 (DQ137790)	221	225	SC_0013	1735381 - 1735818
n	SmD25 (AF202965)	292	292	SC_0041	618665 - 618932
o	Smp_159830	Absent	Present	SC_0223	27722- 27852
p	Smp_159840	Absent	Present	SC_0223	62276- 62409
q	Smp_scaff000131	Present	Present	Chromosome 4	4308079 - 4308184
r	Smp_scaff000154.2	Present	Present	SC_0097	416541 - 416640
s	Smp_scaff000154.5	Present	Absent	SC_0097	273150 - 273279

Tableau 2 : Les marqueurs génétiques utilisés pour le génotypage ciblé. Base de données locale : <http://cbetm.univ-perp.fr/blast/blast.html>, sma_v5.2.chr.fasta (maj : 24-01-12).

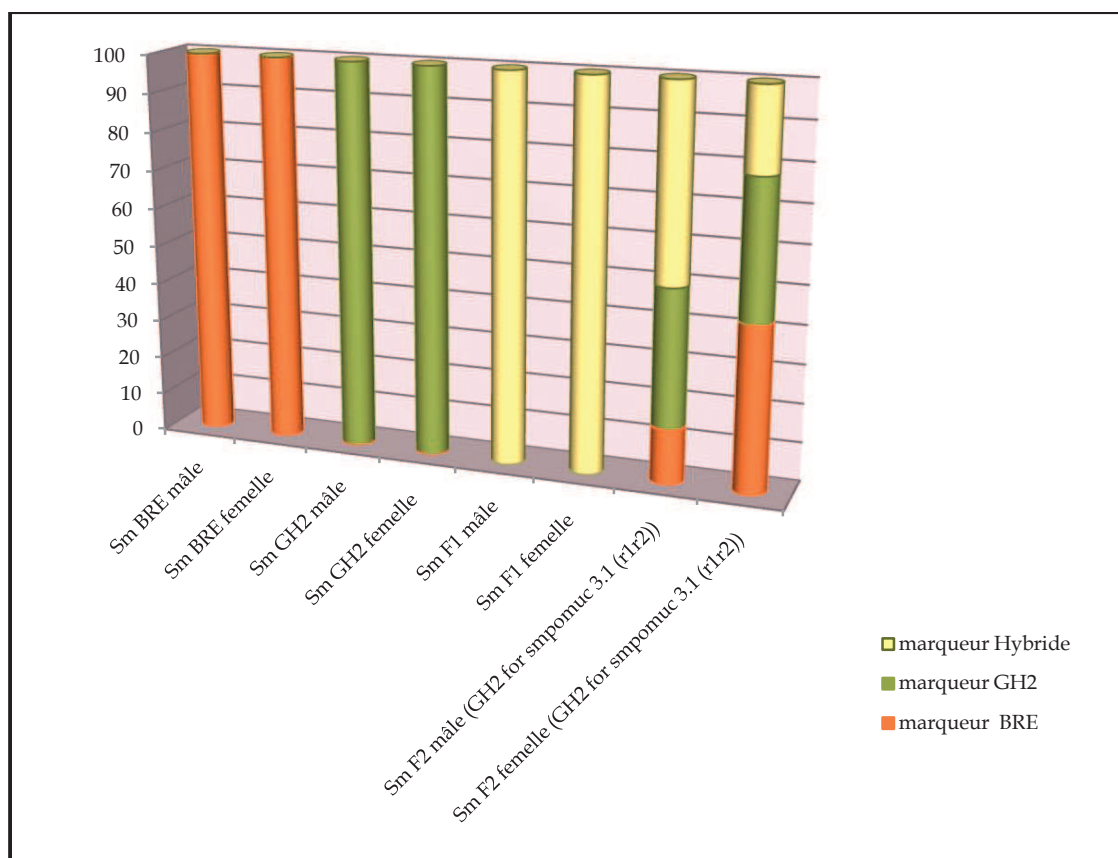


Figure 23 : Génotypage des populations clonales de cercaires. Les populations de cercaires ont été génotypées utilisant des marqueurs spécifiques des deux souches du parasite *Sm* BRE et *Sm* GH2 (marqueur BRE : les deux allèles de BRE, marqueur GH2 : les deux allèles de GH2 et marqueur Hybride : un allèle BRE et un allèle GH2).

II.2.5. Tests statistiques

Les tests statistiques utilisés dans cette étude ont été effectués en collaboration avec Jérôme Boissier. Pour les prévalences, nous avons appliqué le test exact de Fisher. Ce test non paramétrique est préconisé pour comparer des effectifs d'échantillons de petites tailles. Nous avons effectué alors le test sur les effectifs en comparant les prévalences des générations F1, F2 et F3 avec celles de la souche parentale *Sm* GH2 F0. Nous ne les avons pas comparé avec les prévalences de la souche *Sm* BRE F0, parce que cette dernière montre des valeurs optimales de compatibilité.

Afin de tester les différences entre les intensités, un test de Student pour échantillons non appariés a été utilisé. L'intensité correspond à une moyenne obtenue par

différentes valeurs d'intensités d'une même population de miracidia ($27 < n < 50$ mollusques). Pour la même raison que pour les prévalences, nous avons comparé les intensités des générations F1, F2 et F3 avec celles de la souche *Sm* GH2.

Ensuite, pour les différences statistiques en nombre de miracidia par foie et le taux des transcrits (valeur obtenues à partir de tous les résultats de transcription), nous avons utilisés le test ANOVA suivi par le post-hoc test de Dunnett. L'ANOVA a été utilisée pour tester les différences entre les moyennes des différentes valeurs observées de chaque génération. Ensuite, nous nous sommes servis du test du Dunnett pour comparer chacune des générations à une génération contrôle. Dans notre étude, les générations contrôles correspondaient aux souches parentales *Sm* BRE F0 et *Sm* GH2 F0.

Chapitre III

**Mise en évidence des porteurs de l'information
épigénétique chez l'hôte *Biomphalaria glabrata***

Chapitre III : Mise en évidence des porteurs de l'information épigénétique chez l'hôte *Biomphalaria glabrata*

III.1. Objectif scientifique du chapitre III

L'évolution des parasites et des hôtes sont interdépendantes. Dans les interactions hôte / parasite, les espèces interagissent continuellement. Non seulement l'hôte constitue le milieu environnant du parasite, ce dernier constitue aussi un environnement stressant pour l'hôte. D'où, l'hôte *B. glabrata*, comme le parasite *S. mansoni*, devrait exprimer des variants phénotypiques pour s'adapter (Frank 2000).

Nous avons montré dans le chapitre précédant, que les mécanismes épigénétiques pourraient être une source des variants phénotypiques adaptatifs chez le parasite *S. mansoni*. D'après l'hypothèse du « *matching phenotype* » ou bien la concordance des phénotypes appliquée sur notre modèle, la compatibilité nécessite l'expression des variants phénotypiques aussi bien chez le mollusque que chez le parasite (Théron & Coustau 2005). Dans un contexte de coévolution entre le mollusque et le parasite, nous nous sommes interrogés autour de l'implication des mécanismes épigénétiques dans l'expression des variants phénotypiques aussi chez le mollusque.

Knight et ses collaborateurs ont étudié l'influence de la localisation nucléaire d'un locus sur son expression chez le mollusque *B. glabrata* (Arican-Goktas *et al.* 2014). Leurs travaux suggèrent que la distribution de certaines modifications d'histones dans le noyau seraient importantes pour le développement du mollusque (Odoemelam 2009). Ces travaux constituent les premières descriptions de porteurs de l'information épigénétique chez le mollusque *B. glabrata* et par ailleurs, montrent l'importance de cette information dans un contexte d'interaction avec le parasite. A part ce travail, aucune autre marque épigénétique n'avait été étudiée chez le mollusque au début de ma thèse.

L'objectif de ce chapitre était d'initier l'étude des marques épigénétiques chez *B. glabrata*. Nous nous sommes focalisés en premier lieu sur la recherche de la méthylation de l'ADN chez *B. glabrata*. En effet, la méthylation de l'ADN est la

marque épigénétique la plus étudiée chez les invertébrés à l'heure actuelle (Zemach *et al.* 2010; Glastad *et al.* 2011). Par ailleurs, chez un autre mollusque *C. gigas*, la méthylation de l'ADN à l'échelle globale du génome (ou le méthylome) a été récemment caractérisée (Gavery & Roberts 2013).

III.2. Approche méthodologique

1. Quels sont les porteurs d'informations épigénétiques chez *B. glabrata* ?
2. La méthylation de l'ADN existe-t-elle chez *B. glabrata* ? Quel est le taux de 5mC ? 5hmC ? Quel type de méthylation présente cet invertébré ?

Afin d'étudier les porteurs d'informations épigénétiques chez *B. glabrata*, nous avons initié cette étude par l'exploration de la machinerie épigénétique. Cette recherche a été réalisée par une analyse bioinformatique en exploitant les données transcriptomiques du mollusque, disponibles à l'heure actuelle. Ces données ont été générées à partir de trois souches de *B. glabrata* provenant de régions différentes *Bg* BRE (origine Brésil), *Bg* GUA (origine Guadeloupe) et *Bg* BAR (origine Brésil). Au cours de ces analyses, nous nous sommes particulièrement intéressés à la recherche des gènes codant des enzymes impliquées dans la méthylation de l'ADN.

Nos premières investigations et expérimentations ont été consacrées à la mise en évidence de méthylation de l'ADN chez le mollusque. Pour ces expériences Dix individus mollusques *Bg* BRE et dix individus *Bg* GUA ont servi.

Les différentes démarches suivies :

- La présence / absence de méthylation de l'ADN dans le génome de *B. glabrata* a été effectuée sur *Bg* BRE et *Bg* GUA. Nous avons utilisé une technique basée sur la digestion d'ADN par deux différentes enzymes de restriction – isoschizomères *MspI* et *HpaII*, qui reconnaissent et clivent le même site spécifique 5'-CCGG-3'. Cependant, *MspI* est insensible à la méthylation d'ADN et coupe lorsque la cytosine interne est méthylée alors que *HpaII* est sensible à la méthylation.
- La quantification des 5mC et 5hmC dans le génome des mollusques (*Bg* BRE et *Bg* GUA) a été évaluée par Chromatographie en phase liquide- Spectrométrie de masse

(LC/MS), en collaboration avec Michael Reichelt (Max-Planck-Institut fuer Chemische Oekologie, Germany).

- L'étude des ratios CpG/GpC constitue un bon indicateur de la méthylation de l'ADN dans les génomes. Ils nous ont permis de déterminer le type de méthylation chez le mollusque (voir I.3.1.1.2. Méthylation de l'ADN chez les vertébrés *vs* les invertébrés).

Sachant que la méthylation de l'ADN chez les invertébrés a lieu surtout dans les unités transcriptionnelles, nous avons évalué la distribution des ratios CpG/GpC au niveau des séquences exprimées issues du RNA-seq.

- Après avoir démontré l'existence de la méthylation de l'ADN à l'échelle globale du génome chez le mollusque, nous sommes passés à l'échelle ciblée. Pour cette enquête, la séquence choisie correspondait à un élément répétitif nimbus (*Bgl*) du génome de *B. glabrata*, un retrotransposon non-LTR (Genbank, N° accession EF413180). Raghavan et ses collaborateurs ont détecté au moins 100 copies de *Bgl* dans le génome du mollusque, et que cet élément est transcriptionnellement actif dans les différents stades de développement du mollusque (Raghavan *et al.* 2007).

III.3. Publication N°3

5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*

Sara Fneich, Nolwenn Dheilily, Coen Adema, Anne Rognon, Michael Reichelt, Jan Bulla, Christoph Grunau and Céline Cosseau

Parasites & Vectors 2013, 6:167.



RESEARCH

Open Access

5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*

Sara Fneich^{1,2}, Nolwenn Dheilly^{1,2}, Coen Adema⁴, Anne Rognon^{1,2}, Michael Reichelt³, Jan Bulla⁵, Christoph Grunau^{1,2} and Céline Cosseau^{1,2*}

Abstract

Background: *Biomphalaria glabrata* is the mollusc intermediate host for *Schistosoma mansoni*, a digenean flatworm parasite that causes human intestinal schistosomiasis. An estimated 200 million people in 74 countries suffer from schistosomiasis, in terms of morbidity this is the most severe tropical disease after malaria. Epigenetic information informs on the status of gene activity that is heritable, for which changes are reversible and that is not based on the DNA sequence. Epigenetic mechanisms generate variability that provides a source for potentially heritable phenotypic variation and therefore could be involved in the adaptation to environmental constraint. Phenotypic variations are particularly important in host-parasite interactions in which both selective pressure and rate of evolution are high. In this context, epigenetic changes are expected to be major drivers of phenotypic plasticity and co-adaptation between host and parasite. Consequently, with characterization of the genomes of invertebrates that are parasite vectors or intermediate hosts, it is also essential to understand how the epigenetic machinery functions to better decipher the interplay between host and parasite.

Methods: The CpGo/e ratios were used as a proxy to investigate the occurrence of CpG methylation in *B. glabrata* coding regions. The presence of DNA methylation in *B. glabrata* was also confirmed by several experimental approaches: restriction enzymatic digestion with isoschizomers, bisulfite conversion based techniques and LC-MS/MS analysis.

Results: In this work, we report that DNA methylation, which is one of the carriers of epigenetic information, occurs in *B. glabrata*; approximately 2% of cytosine nucleotides are methylated. We describe the methylation machinery of *B. glabrata*. Methylation occurs predominantly at CpG sites, present at high ratios in coding regions of genes associated with housekeeping functions. We also demonstrate by bisulfite treatment that methylation occurs in multiple copies of Nimbus, a transposable element.

Conclusions: This study details DNA methylation for the first time, one of the carriers of epigenetic information in *B. glabrata*. The general characteristics of DNA methylation that we observed in the *B. glabrata* genome conform to what epigenetic studies have reported from other invertebrate species.

Keywords: *Biomphalaria glabrata*, Gastropods, CpG methylation, Epigenetic, *Schistosoma*

* Correspondence: celine.cosseau@univ-perp.fr

¹Université de Perpignan Via Domitia, Perpignan F-66860, France

²UMR 5244 CNRS Ecologie et Evolution des Interactions (2EI), Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, Perpignan Cedex 66860, France

Full list of author information is available at the end of the article

Background

Epigenetic information is information on the status of gene activity that is heritable, for which changes are reversible and that is not based on the DNA sequence. This information can be based on chromatin marking systems such as DNA methylation, histone modifications, and non-coding RNA and the gene location in the nucleus [1-3]. Chromatin exists either as a relaxed structure that is permissive to gene expression and is called euchromatin, or as a condensed structure that is typically silent and is called heterochromatin [1]. Transitions between these different chromatin states may occur in response to environmental signals and result in modification of gene expression, which ultimately, influences phenotypic outcomes without changes in DNA sequence. This environmentally responsive aspect of the chromatin marking system make the Epigenetic Inheritance System a mediator that could allow for rapid adaptive evolution [2,3]. Among the different carriers of the chromatin regulatory system, DNA methylation is one of the best studied. It is conserved in most major eukaryotic groups [4,5], although it has been lost in certain model organisms as listed below [6-8]. The levels, patterns and function of DNA methylation vary among species. In vertebrates and flowering plants, DNA methylation is uniformly distributed in the genome and occurs in transposons (or transposable elements; TE) and repeat elements, in intergenic regions and in transcriptional units including exons and introns. Methylation of promoter sequences has been shown to modulate gene expression, and genetic mobile elements can be silenced through transient methylation [4,5]. Methylation influences important traits such as DNA repair and stability, X chromosome inactivation, inheritable genome imprinting, germ cell pluripotency, as well as development, maintenance and fate of cells.

In invertebrates, the level of DNA methylation varies drastically among species. Some organisms display little or no methylated DNA such as *Tribolium castaneum*, *Drosophila melanogaster* [4] and *Caenorabditis elegans* [8], however, recent investigations have revealed that DNA methylation occurs in many taxa [9-13]. These studies show that genomes of invertebrates are characterized by interspaced regions of methylated and unmethylated DNA. Moreover, intragene (exon and intron) methylation is a general feature of invertebrate organisms [4,5,10,14], while methylation of TE and repetitive elements occur only at moderate levels. Conservation of highly methylated genes is a common feature of invertebrate genome evolution [15], contrasting with plants, for instance, where this is uncommon [16]. Several studies have demonstrated that methylation is likely to occur in genes associated with metabolism and housekeeping functions [11,15], while other studies have suggested that DNA methylation may regulate gene expression by altering the landscape of

splice variants [10,17]. In social insects, important traits have been linked to DNA methylation such as those associated with the social roles in the colony [18-20]. In conclusion, DNA methylation is an important feature of the epigenetic control of the genome. At the same time it appears to be relatively heterogeneous in its taxonomical distribution and it can work through different mechanisms to fulfill even different tasks in different taxa. In a first step towards the study of the contribution of epigenetic mechanisms to the interaction of the human parasite *Schistosoma mansoni* with its intermediate gastropod host, we addressed the question of the type of DNA methylation that occurs in the freshwater snail *Biomphalaria glabrata*. *Schistosoma mansoni* is a digenetic flatworm parasite that causes human intestinal schistosomiasis. An estimated 200 million people in 74 countries suffer from schistosomiasis and it is the most severe tropical parasitic disease in terms of morbidity after malaria [21]. Eggs of the parasite lodge in a definitive host's liver (human or rodent) and accumulate to cause serious disease symptoms. With its feces, the vertebrate host releases many schistosome eggs. When these come into contact with water, free-swimming miracidia hatch and actively seek *B. glabrata* snails (and other species depending on geographic location). After penetration into this intermediate host, the parasite develops into sporocysts that produce cercariae [22]. The latter parasite stage is released into water and can infect the vertebrate host. *Biomphalaria glabrata* belongs to one of the largest invertebrate phyla, the Mollusca, which are lophotrochozoans, a lineage of animal evolution distinct from ecdysoans, represented by model invertebrates such as *Caenorhabditis* and *Drosophila*. We describe here that DNA methylation occurs in *B. glabrata*, with approximately 2% of cytosine nucleotides being methylated. Methylation occurs predominantly at CpG sites and it is found in coding regions of genes associated with housekeeping functions. Methylation was also detected in multiple copies of Nimbus, a transposable element.

Methods

Ethics statement

The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Éducation Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87-848 and 2012201-0008)) for the experimenters. Housing, breeding and animal care followed the national ethical requirements.

Biological material, preparation of genomic DNA

Biomphalaria glabrata were reared in spring water at 25°C, 12 hours light / 12 h dark. Genomic DNA was prepared from soft tissues of the foot of pooled snails. The biological samples for methyl cytosine positive and

negative controls were obtained from *Oncorhynchus mykiss* (see [23]), HeLa cells and from *S. mansoni* adult worms Brazilian strain (*Sm Bre*; see Theron et al. 1997). *S. mansoni* indeed displays no or very low levels of DNA methylation [24-26]. HeLa cells were a generous gift of Albertina De Sario INSERM U827 (IURC). Cells were maintained in DMEM-Glutamax (Gibco) with 10% fetal calf serum and penicillin/streptomycin.

Tissues or HeLa cells were incubated overnight at 55°C with 1 ml of lysis buffer (20 mM Tris/Cl pH 8; 1 mM EDTA; 100 mM NaCl; 0.5% SDS) and 0.3 mg of proteinase K. The samples were then extracted twice with equal volumes of phenol/chloroform, followed by two extractions with equal volumes of chloroform. DNA was precipitated with an equal volume of isopropanol/sodium acetate (3 M, pH5.2) at room temperature. After centrifugation and washing with 1 ml of 70% ethanol, the pellet was dissolved in 100 µl of 1 mM Tris/HCl, pH 8. Finally, DNA was purified using Wizard SV Gel PCR Clean Up System (Promega).

Expressed sequence data resources

The RNA-seq data from 3 different *B. glabrata* isolates (whole snail): a Guadeloupian strain (*Bg Gua*), two Brazilian strains (*Bg Bre* and *Bg Bar*) were provided by Guillaume Mitta (personal communication) for use in this study. Part of the analysis was performed with the RNA-seq data publicly available from snaildb (<http://www.snaildb.org/content/blast>) [27].

Correlation tests

For correlation analysis between the level of transcript expression and the quantity of CpG methylation of these respective transcripts, we performed a correlation test between the transcript rpkm results (Guillaume Mitta, Personal communication) and their respective CpG/e ratio. Both a linear (Pearson's) and non-linear (Spearman's) correlation analysis were performed using the R 2.15.2 (2012 The R Foundation for Statistical Computing ISBN 3-900051-07-0) under a local Galaxy instance [28].

Blast search for molecular machinery for DNA methylation

To investigate the presence of the genes for the molecular machinery for DNA methylation in *B. glabrata*, we performed tblastN against the supercontigs of the preliminary genome assembly version 4.3 (<http://biology.unm.edu/biomphalaria-genome/index.html>). We used as a query, sequences from *Homo sapiens* for which the annotation corresponded to different categories of relevant protein families: DNMTs (DNA Methyl Transferase), Tet (Ten Eleven Translocation enzymes) and MBDs (Methyl-CpG Binding Domain proteins). Expression of each candidate was tested in the different

transcript databases (<http://www.snaildb.org/content/blast>) and the 3 RNA-seq libraries generated in our lab from different isolates of *B. glabrata*.

B. glabrata potential MBD2/3 protein sequence was aligned with the BioEdit software (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>), with the following proteins sequences: *Apis mellifera* MBD2/3 (XP_392422.2), *Homo sapiens* MBD2 (NP_003918.1), *H. sapiens* MBD3 (NP_003917.1), *Hemicentrotus pulcherrimus* MBD2/3 (ACF05485.1), *Crassostrea gigas* MBD2/3 (EKC32831.1), *Ixodes scapularis* MBD2/3 (XP_002407962.1).

Digestion with methylation sensitive restriction enzymes

After DNA extraction, restriction digests with *HpaII* (R0171S, 10000 units/ml, BioLabs) and *MspI* (R0106S, 20000 units/ml, BioLabs) were carried out for two hours at 37°C with 200 ng of genomic DNA and 6 units of enzyme in a final volume of 20 µl. *MspI* and *HpaII* share the same CCGG recognition site but the latter enzyme cannot cut sites with methylated cytosines. DNA fragments were size separated by electrophoresis through a 1% agarose/TBE gel and visualised using ethidium bromide staining.

Liquid chromatography- mass spectrometry (LC-MS) analysis

The levels of 5-methylcytosine (5-mC) and 5-hydroxy methylcytosine (5-hmC) in DNA samples were determined by LC-MS/MS analysis as their deoxyribonucleosides. Pooled genomic DNA of 10 individuals was digested to nucleosides with DNA Degradase Plus (E2020, Zymo research). Fifty units of the enzyme preparation were added to 10 µg of DNA in a total volume of 100 µl. Samples were incubated at 37°C for one hour with low speed rotation. An additional 50 units of enzyme were added to each sample and digestion was continued for one more hour. Complete digestion into nucleosides was verified by visualization of the digest on a 1% agarose gel. Nucleic acid concentration was quantified spectrophotometrically (Nanodrop, Thermo Scientific). Finally, samples were lyophilised and re-suspended in 200 µl of water. A 2 µl aliquot was injected into an LC-MS/MS system for analysis of 5-methyl-2'-deoxycytidine and 5-hydroxymethyl-2'-deoxycytidine, and a 2 µl aliquot of a 1:50 (v:v) diluted sample (dilution in water) was processed for analysis of 2'-deoxycytidine. Chromatography was performed on an Agilent 1200 HPLC system (Agilent Technologies, Boeblingen, Germany) with a Zorbax Eclipse XDB-C18 column (50 × 4.6 mm, 1.8 µm, Agilent Technologies, Germany). Formic acid (0.05%) in water and acetonitrile were employed as mobile phases A and B respectively. The elution profile was: 0-0.5 min, 5% B; 0.5-2 min, 5-100% B in A; 2-3 min 100% B in A; 3-3.1 min 100-5% B and 3.1-5.5 min 5% B. The mobile

phase flow rate was 1.1 ml/min, column temperature was maintained at 25°C. The HPLC was coupled to an API 5000 tandem mass spectrometer (Applied Biosystems, Darmstadt, Germany) equipped with a Turbospray ion source operated in positive ionization mode. The instrument parameters were optimized by infusion experiments with pure standards: 2'-deoxycytidine (Acros Organics, New Jersey, USA); 5-methyl-2'-deoxycytidine (ABCR, Karlsruhe, Germany); 5-hydroxymethyl-2'-deoxycytidine (Berry & Associates, Ann Arbor, USA). The ionspray voltage was maintained at 5500 eV. The turbo gas temperature was set at 700°C. Nebulizing gas was set at 70 psi, curtain gas at 25 psi, heating gas at 60 psi and collision gas at 6 psi. Multiple reaction monitoring (MRM) was used to monitor analyte parent ion → product ion: m/z 228 → 112 (collision energy (CE) 17 V; declustering potential (DP) 45 V) for 2'-deoxycytidine; m/z 242 → 126 (CE 15 V; DP 51 V) for 5-methyl-2'-deoxycytidine; m/z 258 → 142 (CE 15 V; DP 41 V) for 5-hydroxymethyl-2'-deoxycytidine. Both Q1 and Q3 quadrupoles were maintained at unit resolution. Analyst 1.5 software (Applied Biosystems, Darmstadt, Germany) was used for data acquisition and processing. Linearity in ionization efficiencies were verified by analyzing dilution series of authentic standards. External calibration curves for 2'-deoxycytidine, 5-methyl-2'-deoxycytidine, and 5-hydroxymethyl-2'-deoxycytidine were used for quantification.

Bisulfite treatment, PCR amplification, COBRA and sequencing

Bisulphite sequencing allows identifying DNA methylation at single nucleotide resolution. The treatment of DNA with bisulfite converts all non-methylated cytosines into deoxy-uracil while the 5-methylcytosines (5-mC) and 5-hydroxymethyl-cytosines (5-hmC) remain intact. To address the question whether repetitive elements are methylated and to confirm the occurrence of CpG methylation in transcripts in *B. glabrata*, we analysed bisulfite treated gDNA by PCR targeting the non-LTR retrotransposon nimbus (*Bgl*). *Bgl* is transcriptionally active in different stages of the mollusc development, and at least 100 copies were detected in the *B. glabrata* genome [29,30]. Bisulfite treatment was carried out as described previously [31,32] (<http://methdb.univ-perp.fr/epievo/>). Briefly, 300 ng DNA was denaturated with 3 M NaOH, treated with a solution of sodium-bisulfite and hydroquinone at pH 5 in the dark for 4 hours at 55°C; desalted (Amicon Ultra column, UFC510024 Millipore), desulfonated by adding 350 µl of 0.1 M NaOH to the DNA in the Microcon column, and finally dissolved in 10 mM Tris/Cl pH 8. The genomic sequence of the transposable element Nimbus *Bgl* was obtained from Genbank (acc.nr. EF413180, [29]). Primer pairs (Table 1) were designed using MethPrimer [33]. Six

ng of treated DNA were used as a template for PCR amplification in 25 µl using Go Taq DNA polymerase (Promega) at 94°C for 2 min, 5 cycles of 94°C for 1 min, 50°C for 2 min and 72°C for 3 min; followed by 25 cycles of 94°C for 30 sec, 50°C for 2 min and 72°C for 1:30 min and finally 72°C for 10 min. PCR products were separated by electrophoresis through 1% agarose gels. The degree of methylation was estimated by COmbined Bisulfite Restriction Assay (COBRA) [34]. Suitable restriction enzymes were selected with Snake Charmer (http://insilico.ehu.es/restriction/two_seq/snake_charmer.html) and used to digest each of the PCR products by three different enzymes: *AccI* (NEB); *TaiI* (Fermentas); *TaqI* (Promega) using 15 units to digest 100 ng in each reaction. Digestion fragments were separated on a 2% agarose gel. For high resolution analysis, 1 µl of each PCR product was cloned into pCR4 (TOPO TA Cloning kit, Invitrogen) and 11 inserts derived from *Bg Bre* and *Bg Gua* DNA each were sequenced with vector specific primers (M13F: CTGGCC GTCGTTTTAC and M13R: CAGGAAACAGCTATGAC) using the GATC Biotech facilities (<http://www.gatc-biotech.com/fr/entreprise/gatc.html>). Chromatograms were edited using Sequencher Software, and a total of 14 sequences (combined from *Bg Bre* and *Bg Gua*) were retained for final analysis. These experimentally obtained sequences were aligned with the genomic sequence from GenBank (Bioedit) to visualise the sites of methylated cytosine. MethTools 2.0 software [35] was used to generate a graphical view of the *Bgl* region 2 containing the methylated sites.

CpG observed/expected ratio analysis

The CpG observed/expected ratio (CpGo/e) analysis was performed on cDNA sequences (≥500 bp) of 4 different RNA-Seq assemblies: 1 public dataset (<http://www.snaildb.org/webfm>), representing 12579 uniseq from the *Bg Lilles* strain [27] and 4 libraries generated in our laboratory from different geographical isolates of *B. glabrata* (42824 uniseq - strain: *Bg Bre*, 22088 uniseq - strain: *Bg Gua* and 16855 uniseq - strain: *Bg Bar*). The CpGo/e ratio was used as a proxy to estimate the intragene DNA methylation content of *B. glabrata* as previously described in other invertebrate species [11,36,37]. CpGo/e was calculated from 5000 uniseq randomly chosen in each RNA-seq library. CpGo/e was calculated for the coding strand only, using the following equation where l is the number of nucleotides [11].

$$CpGo/e = \frac{\text{number.of.CpG}}{\text{number.of.C} \times \text{number.of.G}} \times \frac{l^2}{l-1}$$

To discern two subgroups with high and low CpGo/e values, statistical analyses were carried out using the software R 2.15.2 (www.r-project.org). The clustering algorithm employed utilizes the Mclust procedure from the package mclust, version 4.0 [38,39]. Akaike information

Table 1 Primers used in this study

Region	Primer names	Sequence and 5' position*		Amplicon length
		Forward	Reverse	
Region 1	Nimbus 1	5'-TTTTATGAGGTGTTTAAAGTGTAGG-3' (5' position: 1132)	5'-AAAAATTCCTTTATTCCAATAAC-3' (5' position: 1917)	785 bp
Region 2	Nimbus 2	5'-TTGGATGTTAAATTTTGTAGAA-3' (5' position: 2986)	5'-AAAAAATATCCCTTAAACCCATAA-3' (5' position: 3871)	785 bp
Region 3	Nimbus 3	5'-ATTTTAGGGAATTGTAGGAGAGTTA-3' (5' position: 4802)	5'-CTTATCAAACCTTAAATATAAAC-3' (5' position: 5373)	571 bp

*position is given on the genbank sequence: EF413180.

criterion (AIC) and Bayesian information criterion (BIC) were used as selection criteria to establish if a simple Gaussian or a mixture of two Gaussian distributions fit best to empirical distribution of CpGo/e values of *B. glabrata* (Gua strain). The 95% confidence intervals for the mean values were determined by non-parametric bootstrap with 1000 replications.

Functional analysis on low versus high methylated genes

Transcripts from the *Bg* Gua transcriptomic library were chosen for further functional analysis. Functional annotation was performed on the *Bg* Gua data set (22088 unisequences >=500 pb) using Blast2GO version 2.4.2 [40]. The pipeline annotation and parameters used were as follows: i) an initial annotation with BLASTX (against the nonredundant NCBI database; e-value at 1.10^{-6}); ii) a protein domain search using InterProscan; iii) an enzyme annotation using the Kyoto Encyclopedia of Genes and Genomes (KEGG); and iv) assignment of a Gene Ontology term (GO; <http://www.geneontology.org/>).

Bg Gua cDNA sequences (>= 500 pb) were divided into 11 cDNA subsets based on their CpGo/e ratio (0-0.1: 1283 uniseq, 0.1-0.2: 3878 uniseq, 0.2-0.3: 4345 uniseq, 0.3-0.4: 3106 uniseq, 0.4-0.5: 2165 uniseq, 0.5-0.6: 1669 uniseq, 0.6-0.7: 1665 uniseq, 0.7-0.8: 1425 uniseq, 0.8-0.9: 989 uniseq, 0.9-1: 788 uniseq, >1: 775 uniseq). The cDNA set from each subcategory was independently tested (exact Fisher test, p-value < 0.05, two-tailed) for enrichment in functional categories (GO terms) against the complete *Bg* Gua transcriptome set (>= 500 pb, 22088 uniseqs). Over-represented or under-represented GO terms were classified in 3 subcategories independently by 5 experienced biologists: housekeeping GO, non informative GO, inducible GO. GO terms indicating a localization (for example: nucleolus) or a general function (for example: ATP binding) that could either be involved in housekeeping functions or inducible functions were considered as non informative.

Results

Essential elements of the DNA methylation machinery are present in *B. glabrata*

The DNA methyltransferase (DNMT) proteins are a family of enzymes that catalyse the transfer of a methyl group

to the carbon atom number 5 of the cytosin moiety. Three classes of DNMTs have been characterized in mammals. DNMT1 is responsible for the maintenance of methylation and preferentially methylates hemimethylated CpG di-nucleotides. DNMT2 has strong sequence similarities with 5-methylcytosine methyltransferases but has only been shown to methylate tRNA and not DNA [41]. Therefore, its role as a DNA methyl transferase per se remains enigmatic. DNMT3 is responsible for *de novo* CpG methylation and occurs in all mammalian and some invertebrate species. We investigated the occurrence of these 3 enzymes in *B. glabrata* by BLAST search against the preliminary assembly v4.3 of the *B. glabrata* genome. The supercontigs 19819.1, 15589.2 and 15589.1 contain protein-encoding sequences similar to the carboxy terminal, central and amino terminal part of the human DNMT1 sequence, respectively, suggesting that one gene encoding DNMT1 is present in the *B. glabrata* genome (Additional file 1). The supercontig 6274.3 contains genomic sequences encoding a protein that is conserved with the DNMT2 query sequence. Partial transcripts corresponding to both the DNMT1 and DNMT2 candidates were detected in the *B. glabrata* RNAseq libraries. Therefore, the *B. glabrata* genome includes unique genes encoding DNMT1 and DNMT2 homologs, and these are expressed as transcripts. No sequence matches the DNMT3 query sequence, either in the genome draft or in the RNA seq libraries.

The Ten Eleven Translocation enzymes (Tet1, 2, 3) are a family of oxygenases that catalyse the conversion of 5-methylcytosine (5-mC) into 5-hydroxymethylcytosine 5-hmC in DNA. 5-hmC has long been believed to be an intermediate in the DNA demethylation pathway but recent evidence indicates that its role in epigenetics may have been underrated so far [42,43]. We performed tBlastN against the preliminary genome assembly and found that two *B. glabrata* supercontigs (Contig18827.1 and Contig21.52) contain sequences that match the oxygenase domains present in tet enzymes but that lack the amino terminal part of the tet human orthologs (Additional file 1). The % of coverage of the query sequence is only 50%. No transcripts with similarity to tet enzymes were detected in the RNA-seq libraries. Therefore, it remains an open question

whether true “tet enzymes” or unrelated proteins with oxygenase activity occur in *B. glabrata*.

The MBD2 and MBD3 (Methyl-CpG Binding Domain proteins) query sequences match with the same unique target in the *B. glabrata* genome, which is located on contig 7228 (Additional file 1). This would be consistent with studies that report that other invertebrate species differ from vertebrates in that they possess only one gene that combines the function of MBD2 and MBD3 proteins, whereas homologs of MBD1 and 4 are not detected [44]. All four RNA-seq libraries included full length transcripts of this sequence, therefore the MBD2/3 gene could be reconstituted and sequence alignments indicate closed conservation with MBD2/3 sequences from other invertebrate species (Additional file 2).

The available data indicate that critical elements of the machinery for DNA methylation are present, suggesting that *B. glabrata* DNA could be methylated.

Genomic DNA of *B. glabrata* is only partially digested with methylation sensitive restriction enzymes

Digestion of *B. glabrata* gDNA with methylation sensitive *HpaII* (restriction site 5'-CCGG-3') revealed an undigested high molecular weight fraction as compared to the continuous smear of digestion products produced by its isoschizomer *MspII* that is insensitive to methylation (Figure 1). In contrast, gDNA from *S. mansoni* that contains very low levels of methyl cytosine [24-26], was digested to completion with both *MspII* or *HpaII*: no undigested high molecular weight fraction remained. The DNA from *O. mykiss* and HeLa cells (positive controls) displayed a more pronounced resistance to *HpaII* digestion, consistent with the relatively high levels of CpG methylation observed in these organisms [23]. The resistance to *HpaII* digestion provides an indication for methylation of CpG in restriction sites present in the genomic DNA of *B. glabrata*. The gel aspects of the restriction digests were similar for two different strains of *B. glabrata* species; indicating no obvious differences in CpG methylation profiles for these two strains.

Two percent of total cytosines are methylated in *B. glabrata*

For a precise measurement of the amount of methylated and hydroxymethylated cytosines, genomic DNA of *Bg* Gua and *Bg* Bre was digested into nucleosides and the occurrence of mC and hmC was analysed by LC-MS/MS. The results show that in the *B. glabrata* genome, about 2% of total cytosines are methylated and 0.0009% of total cytosines are hydroxymethylated (Table 2). The control, consisting of an unmethylated PCR product tested negative for either type of modified cytosines.

B. glabrata transcripts are divided in low and high methylated genes

We calculated the empirical distribution of CpGo/e ratios for the transcripts from the 3 RNAseq libraries issued from different geographical areas and from the recently published library from the strain *Bg* Lille [27]. The profile from all the libraries was the same (data not shown). A mixture of two Gaussian distributions as well as a simple Gaussian distribution were fitted to the CpGo/e ratio data from the *Bg* Gua library. The model selection criteria AIC and BIC both indicate a clear preference for the mixture model. The AIC values equal 1243 and 386 for the simple Gaussian and the mixture model respectively and the BIC values equal 1256 and 419 for the simple Gaussian and the mixture model respectively. Figure 2 displays a histogram of *Bg* Gua CpGo/e ratios with the fitted mixture distribution. The estimated mean values of the two components are 0.209 and 0.616. Their 95% confidence intervals equal [0.201, 0.218] and [0.604, 0.628], respectively. The two clearly distinct means with non-overlapping confidence intervals underline the bi-modality, indicating the presence of two sub-populations in the sample: presumably high-methylated genes and presumably low methylated-genes. The proportions of these two potential sub-populations can be inferred a posteriori as by-products of the parameter estimation procedure: the corresponding values equal 33.3% and 66.7% for the component with lower and higher means, respectively.

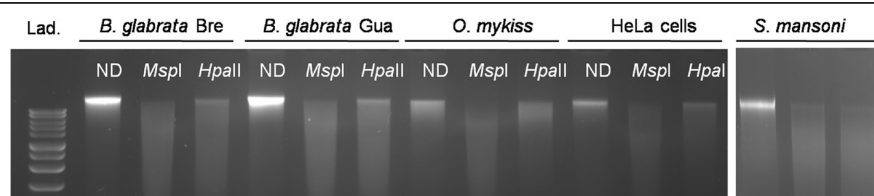


Figure 1 Methylation sensitive restriction assay. Genomic DNA of *B. glabrata*: Brazilian strain (*B. glabrata* Bre) and Guadeloupian strain (*B. glabrata* Gua), of positive controls: *O. mykiss* and HeLa cells and of negative controls: *S. mansoni* (miracidia) was digested with the *MspI* enzyme, its methylation sensitive isoschizomer *HpaII* or was not digested (ND). 200 ng of DNA were loaded per well. Digestion profile was observed on an ethidium bromide stained 1% agarose gel. Lad: 1 kb DNA ladder (promega), lowest band 1500 bp, highest band 10000 bp.

Table 2 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) levels in *B. glabrata* strains as determined by LC/MS

<i>B. glabrata</i> strains	% 5-mC	% 5-hmC
Guadeloupian strain	2.1	0.00091
Brazilian strain	2.07	0.0009

Gene methylation categories are correlated with functions of the gene products but not with transcription level

We retrieved the lists of transcripts that display CpGo/e ratios between 0-0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4, 0.4-0.5, 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, 0.9-1 and >1. Each of these 11 subcategories was submitted to a Fisher exact test in the blast2Go web-based interfaced to test for enrichment in functional categories in each of these subclasses. GO categories (Gene Ontology categories), that were statistically enriched were then classified as housekeeping, non informative or inducible (see Additional file 3 for these assignments). The GO categories over-represented in the subset of CpGo/e ratios from 0.1 to 0.4 were more enriched in housekeeping functions, whereas subsets of CpGo/e ratios from 0.4 to >1 were more enriched in inducible categories (Table 3). We observed a transition for categories in the subset of CpGo/e ratio from 0.4 to 0.6. Expectedly, the GO categories under-represented in the subset of CpGo/e ratio from 0.4 to >1 were more enriched in housekeeping categories (Table 3). Therefore, the high methylated gene category is more enriched in genes encoding housekeeping functions as previously shown for other invertebrates [11].

We tested whether *Biomphalaria* DNA methylation is related to transcript expression. CpGo/e ratio versus rpkms correlation analysis did not give a significant p-value for both Pearson's and Spearman's correlation analysis. Therefore, we did not find a significant genome-wide correlation between transcript CpG methylation and gene expression.

Many copies of the *B. glabrata* non-LTR retrotransposon nimbus (*Bgl*) are highly methylated

To further analyze whether DNA methylation also occurs in intergenic regions and in particular in repetitive sequences as represented by a retrotransposon like *Bgl* Nimbus, we used bisulfite conversion based techniques. After bisulfite treatment of genomic DNA from *B. glabrata*, three different regions of the transcriptionally active non-LTR retrotransposon nimbus (*Bgl*) were amplified by PCR from each *B. glabrata* strain. Then, digestion of this PCR product with restriction enzymes that target sites containing cytosine (*Acil*, *Tail* or *TaqI*), indicated methylation in *Bgl* region 2. It indicates that this region has methylated cytosines. For high resolution mapping of the methylated sites, 14 amplicon sequences of region 2 (mixed *Bg* Gua and *Bg* Bre) were cloned and sequenced. Alignments of these sequences allowed us to detect that 5 out of the 14 sequences contained methylated CpG sites (Figure 3). A total of 22 methylated CpG sites were identified among the 5 clones (Figure 3).

Discussion

DNA methylation is a covalent modification that consists of adding a methyl moiety from a folate donor to

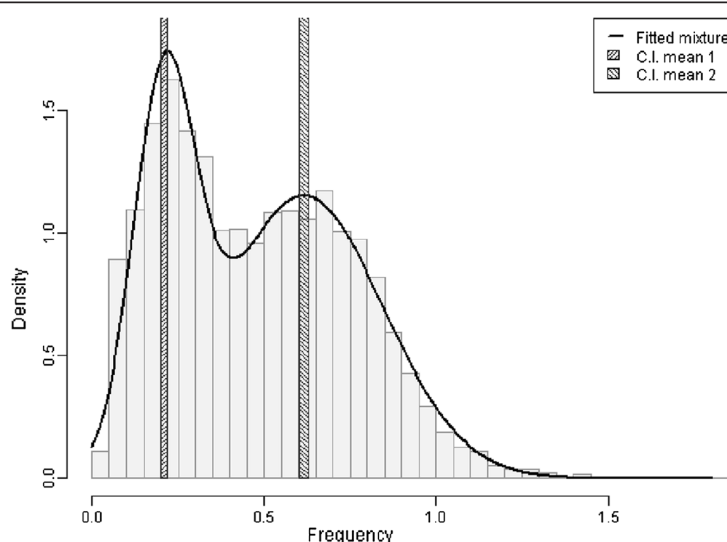


Figure 2 Histogram of CpGo/e ratio in *B. glabrata* transcripts. CpGo/e ratio was measured as a proxy to estimate the CpG methylation in transcripts from RNA-seq libraries from *B. glabrata* guadeloupian strain (*Bg* Gua). X axis: CpGo/e ratio, Y-axis density of transcripts. The figure displays a histogram of *Bg* Gua CpGo/e ratios with a fitted mixture distribution. The grey shaded bars represent 95% confidence intervals for the two mean values. The estimated mean values of the two components are 0.209 and 0.616.

Table 3 Functional analysis on transcripts depending on their CpGo/e ratio

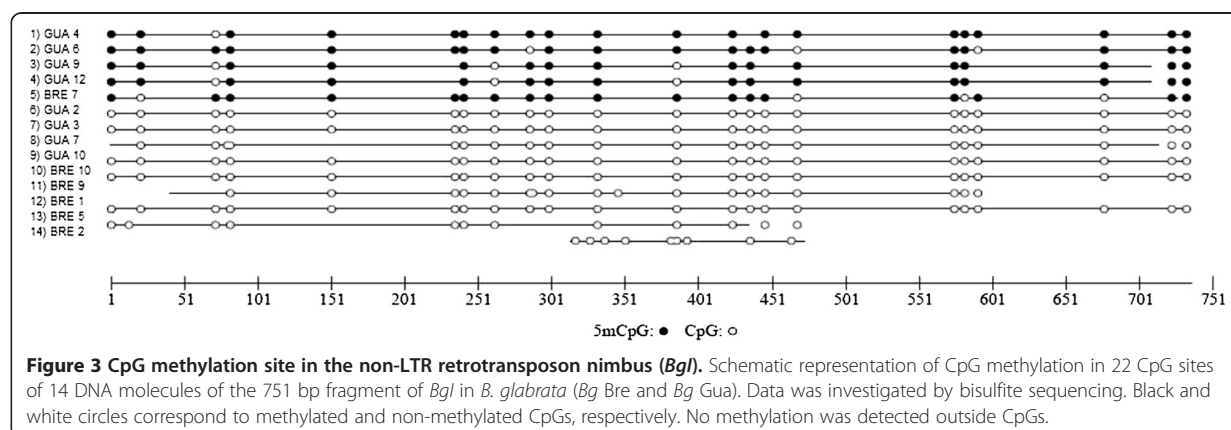
CpGo/e ratio	Over-represented GO terms			Under-represented GO terms		
	Housekeeping functions	Inducible functions	Non informative GO terms	Housekeeping functions	Inducible functions	Non informative GO terms
0-0.1		No enriched categories			No enriched categories	
0.1-0.2	5	2	5	4	2	3
0.2-0.3	14	1	9	1	3	2
0.3-0.4	12	0	8	0	1	1
0.4-0.5		No enriched categories			No enriched categories	
0.5-0.6	3	1	3	2	1	5
0.6-0.7	0	3	2		No enriched categories	
0.7-0.8	1	8	2	14	1	13
0.8-0.9	0	0	2	7	1	6
0.9-1	0	1	3	2	0	3
>1	7	12	4	1	0	12

Bg Gua cDNA (> 500 pb) were divided into 11 cDNA subsets based on their CpGo/e ratio (Left column: 0-0.1 to >1). The cDNA set from each subcategory was independently tested for enrichment in functional categories (GO terms) against the complete *Bg* Gua transcriptome set. Over-represented or under-represented GO terms were classified in 3 subcategories: housekeeping GO, non informative GO, inducible GO (See Additional file 3 for assignment in these subcategories). The numbers represent the total number of GO terms that belong to one of these subcategories.

the carbon atom number 5 of cytosine. This modification is an important epigenetic determinant involved in control of gene expression, maintenance of DNA integrity and stability, one that can be inherited through cell division. While the role of this epigenetic modification has been intensively studied in vertebrate species, plants and invertebrates from the ecdysoan lineage [36], literature reports on molecular epigenetic studies in lophotrochozoan protostome invertebrates such as molluscs are rare [11], especially because of the lack of genomic molecular resources for this phyla. Our study provides the first evidence of DNA methylation in the genome of *B. glabrata*, a gastropod mollusc and further expands insights of DNA methylation in the mollusc phylum, previously only reported for the class of bivalve molluscs: oyster *Crassostrea gigas* [11] and the scallop *Chlamys farreri* [45]. We estimate the percentage of methylated cytosines in the genome of *B. glabrata* to be

roughly 2% by LC-MS analysis of nucleosides. This rate of DNA methylation is rather high for an invertebrate whose general percentage of cytosine being methylated ranks from 0.1 to 4% [46].

We investigated the presence of molecular components that make up the DNA methylation machinery in *B. glabrata*. DNMT1, involved in maintenance of CpG methylation, and DNMT2, characterised as a tRNA methyl transferase, are present in *B. glabrata*, while DNMT3, involved in *de novo* methylation, seems to be lacking in this organism. This result provides another example of organisms that display a functional methylation system in a genome where DNMT3 is lacking. *Schistocerca gregaria*, *Locusta migratoria*, *Bombyx mori* and *Pediculus humanus* are other invertebrate species for which no DNMT3 has been detected, although they display functional DNA methylation machinery [36,37,47,48]. It should be noted that the absence of DNMT3 is not a general feature in



molluscs as DNMT3 is present in *C. Gigas* [11]. For the organisms lacking DNMT3, the question remains open as to what the machinery is for the *de novo* methylation. DNMT1 is a potential candidate; it has been reported to function in the role of a *de novo* methyltransferase in human cancer cells [49]. Other candidates involved in the DNA methylation machinery of *B. glabrata* belong to the family of Methyl-CpG Binding Domain proteins (MBDs). MBDs specifically bind methyl-CpG and interact with the chromatin remodeling complex, which results in other local epigenetic modifications. Four classes of MBDs have been described (MBD1 to 4) [44]. Consistent with observations from other invertebrate species, a single gene encoding for a protein that may combine the function of MBD2 and MBD3 proteins occurs in *B. glabrata* and MBD1 and 4 are not detected [44].

Different types of DNA methylation have been found in invertebrates. Some species dominantly display methylation in the context of CpG, such as *C. intestinalis*, *S. gregaria*, *B. mori*, *M. extrudentata* and *A. mellifera* [5,9,14,47,48,50], while other species such as *P. lilacinus* and *D. melanogaster*; *Pogonomyrmex barbatus* display a high proportion of methylation in non-CpG methylation sites [10,13,51]. The bisulfite sequencing analysis performed in our study identified only methylation of CpG sites, suggesting that this is the dominant type of DNA methylation in *B. glabrata*. The bisulfite sequencing analysis targeted the non LTR-retrotransposon nimbus (*Bgl*) [29], a Mobile Genetic Element (MGE) that is constitutively expressed in naïve *B. glabrata* and its transcription is enhanced upon stress conditions (heat shock and *S. mansoni* infection) [30]. An estimated 100 copies of *Bgl* are present in the *B. glabrata* genome [29]. We detected 22 sites of CpG methylation in a 752 bp region of *Bgl*, and observed variation in the methylation status of the particular CpG sites among the clones that were analysed. Variability of the sources of the *Bgl* sequences, different types of cells, developmental stages, individuals and genomic location of *Bgl* copies, may be responsible for this variation in methylation of specific CpG sites. Patterns of MGEs methylation vary considerably in invertebrates. In some species, MGEs methylation is modest and not a preferred target (*C. intestinalis*, *Drosophila*, *C. farreri*, *A. mellifera* and *B. mori* [4,5,14,48,50,52]), while in other invertebrates repetitive DNA seems to be preferentially methylated (*L. migratoria*, *Medauroidea extrudentata* and *S. gregaria* [9,37,47]). These discrepancies question the idea of a conserved role in invertebrates for methylation in suppressing transcription of transposons as has been demonstrated in mammals and flowering plants [46].

Methylated cytosines are hypermutable because they are spontaneously transformed into thymine by deamination [53,54]. This spontaneous mutation is not recognised by the DNA repair machinery and leaves an evolutionary

signature in the genome of organisms that display CpG methylation or have displayed CpG methylation in the past and can be identified through *in silico* analysis. Consistently, methylated regions of DNA have a globally reduced frequency of CpG dinucleotides compared with the expected frequency. Consequently, there is a negative correlation between the CpGo/e ratio and the degree of methylation, as has been validated in several insect species [15]. In *B. glabrata*, we evaluated the distribution of CpGo/e ratios among expressed sequences from RNA-seq libraries. We clearly observed that most of the sequences display a CpGo/e ratio below 1, consistent with CpG methylation in the units of transcription (transcribed genes). The presence of intragenic methylation in *B. glabrata* is in line with observations reported from other invertebrate species [4,5,14,15]. Furthermore, we observed a bimodal distribution of the normalised CpG ratio, with two peaks, at 0.21 and 0.69 respectively. This bimodal distribution was also observed previously in several other invertebrate species [15,36,55] and indicates the presence of two categories of genes: Low CpGo/e transcripts corresponding to high-methylated genes and high- CpGo/e transcripts corresponding to low methylated-genes. Functional annotation of these genes shows that transcripts predicted to be highly methylated are associated with housekeeping functions and genes predicted to be poorly methylated are associated with inducible functions. This correlation has already been reported in other invertebrate species such as *C. gigas* [11] and the *A. mellifera* [56].

Epigenetic mechanisms shape the expression of the genome during development. In addition, they provide a source for potentially heritable phenotypic variation and could, therefore, be involved in the adaptation to environmental constraints such as parasitic infections. Consequently, it is indispensable to understand the genome of parasite vectors and intermediate hosts but also how the epigenetic machinery functions. This study provides the first report of a precise estimate for the amount of methylated DNA, one of the carriers of epigenetic information, in *B. glabrata* and we deliver insights into its distribution along the genome. We hope to pave the way for a more thorough analysis of the role of DNA methylation in the susceptibility towards infection by parasites such as *S. mansoni*.

Additional files

Additional file 1: Blast analysis on methylation machinery candidates.

Additional file 2: Sequence alignments of MBD2/3 protein families.

B. glabrata MBD2/3 protein sequence was aligned with the BioEdit software (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) with the following protein sequences: *Apis mellifera* MBD2/3 (XP_392422.2), *Homo sapiens* MBD2 (NP_003918.1), *H. sapiens* MBD3 (NP_003917.1), *Hemicentrotus pulcherrimus* MBD2/3 (ACF05485.1), *Crassostrea gigas* MBD2/3 (EKC32831.1), *Ixodes scapularis* MBD2/3 (XP_002407962.1).

Additional file 3: Classification of GO functional categories in housekeeping, inducible or non informative subcategories. *Bg* cDNA (> = 500 pb) were divided into 11 cDNA subsets based on their CpGo/e ratio. Each subset was tested for enrichment in functional category using the Blast2GO software. Over-represented or under-represented GO terms (p value < 0.05) were classified into 3 subcategories: housekeeping GO (underlined in red), non informative GO (not underlined), inducible GO (underlined in yellow).

Abbreviations

LC-MS: Liquid chromatography-mass spectrometry; HPLC: High performance liquid chromatography; MRM: Multiple reaction monitoring; CE: Collision energy; 5-mC: 5-methylcytosine; 5-hmC: 5-hydroxymethylcytosine; MBD: Methyl-CpG binding domain protein; Tet: Ten eleven translocation enzymes; DNMT: DNA methyltransferase; GO: Gene ontology; TBE: Tris-borate-EDTA; CpGo/e: CpG observed/expected ratio; *Bgl*: Non-LTR retrotransposon nimbus; LTR: Long terminal repeats; MGE: Mobile genetic element; AIC: Akaike information criterion; BIC: Bayesian information criterion.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SF performed the experiments, analysed the data and contributed to preparation of the reagents, the materials and the analysis tools. She also took part in the writing of the manuscript. ND contributed to preparation of RNA-seq libraries and participated in the data analysis. CA took part in the writing of the manuscript. AR contributed to the preparation of the reagents and the biological samples. MR performed the LC-MS experiments. JB designed, performed and described the statistical test that was used to test for bimodal distribution of observed/expected CpG ratios. CG conceived and designed the experiments, analysed the data, contributed to preparation of the analysis tools and wrote the manuscript. CC conceived and designed the experiments, performed the experiments, analysed the data, and wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

The authors are indebted to Bernard Dejean and Nathalie Arancibia for providing valuable technical support. The authors acknowledge Dr Benjamin Gourbal, Guillaume Mita, Julien Portela and Richard Galinier for their support through the generation of the RNA-Seq libraries. The work received funding from project ANR-2010-BLAN-1720-01 (EPIGEVOL) and ANR-2010-BLAN-1702 (BODYGUARD) of the French National Agency for Research through a doctoral fellowship. CMA acknowledges support from The National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM103452.

Author details

¹Université de Perpignan Via Domitia, Perpignan F-66860, France. ²UMR 5244 CNRS Ecologie et Evolution des Interactions (2EI), Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, Perpignan Cedex 66860, France. ³Abt. Biochemie, Max-Planck-Institut für Chemische Ökologie, Jena D-07743, Germany. ⁴Center for Evolutionary and theoretical Immunology, Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA. ⁵LMNO, CNRS UMR 6139, Université de Caen, Caen Cedex 14032, France.

Received: 1 March 2013 Accepted: 27 May 2013
Published: 6 June 2013

References

- Luijsterburg MS, White MF, van Driel R, Dame RT: **The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes.** *Crit Rev Biochem Mol Biol* 2008, **43**(6):393-418.
- Turner BM: **Epigenetic responses to environmental change and their evolutionary implications.** *Philos Trans R Soc Lond B Biol Sci* 2009, **364**(1534):3403-3418.
- Jablonka E, Ouborny B, Molnar I, Kisdi E, Hofbauer J, Czarán T: **The adaptive advantage of phenotypic memory in changing environments.** *Philos Trans R Soc Lond B Biol Sci* 1995, **350**(1332):133-141.
- Zemach A, McDaniel IE, Silva P, Zilberman D: **Genome-wide evolutionary analysis of eukaryotic DNA methylation.** *Science* 2010, **328**(5980):916-919.
- Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al: **Conservation and divergence of methylation patterning in plants and animals.** *Proc Natl Acad Sci USA* 2010, **107**(19):8689-8694.
- Rae PM, Steele RE: **Absence of cytosine methylation at C-C-G-G and G-C-G-C sites in the rDNA coding regions and intervening sequences of *Drosophila* and the rDNA of other insects.** *Nucleic Acids Res* 1979, **6**(9):2987-2995.
- Urieli-Shoval S, Gruenbaum Y, Sedat J, Razin A: **The absence of detectable methylated bases in *Drosophila melanogaster* DNA.** *FEBS Lett* 1982, **146**(1):148-152.
- Simpson VJ, Johnson TE, Hammen RF: ***Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging.** *Nucleic Acids Res* 1986, **14**(16):6711-6719.
- Krauss V, Eisenhardt C, Unger T: **The genome of the stick insect *Medauroidea extradentata* is strongly methylated within genes and repetitive DNA.** *PLoS One* 2009, **4**(9):e7223.
- Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R: **The honey bee epigenomes: differential methylation of brain DNA in queens and workers.** *PLoS Biol* 2010, **8**(11):e1000506.
- Gavery MR, Roberts SB: **DNA methylation patterns provide insight into epigenetic regulation in the pacific oyster (*Crassostrea gigas*).** *BMC Genomics* 2010, **11**:483.
- Walsh TK, Brisson JA, Robertson HM, Gordon K, Jaubert-Possamai S, Tagu D, Edwards OR: **A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum*.** *Insect Mol Biol* 2010, **19**(Suppl 2):215-228.
- Smith CR, Mutti NS, Jasper WC, Naidu A, Smith CD, Gadau J: **Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination.** *PLoS One* 2012, **7**(8):e42433.
- Suzuki MM, Kerr AR, De Sousa D, Bird A: **CpG methylation is targeted to transcription units in an invertebrate genome.** *Genome Res* 2007, **17**(5):625-631.
- Sarda S, Zeng J, Hunt BG, Yi SV: **The evolution of invertebrate gene body methylation.** *Mol Biol Evol* 2012, **29**(8):1907-1916.
- Takuno S, Gaut BS: **Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly.** *Mol Biol Evol* 2012, **29**(1):219-227.
- Flores KB, Wolschin F, Allen AN, Comevaux JJ, Huentelman M, Amdam GV: **Genome-wide association between DNA methylation and alternative splicing in an invertebrate.** *BMC Genomics* 2012, **13**(1):480.
- Kucharski R, Maleszka J, Foret S, Maleszka R: **Nutritional control of reproductive status in honeybees via DNA methylation.** *Science* 2008, **319**(5871):1827-1830.
- Lockett GA, Helliwell P, Maleszka R: **Involvement of DNA methylation in memory processing in the honey bee.** *Neuroreport* 2010, **21**(12):812-816.
- Herb BR, Wolschin F, Hansen KD, Aryee MJ, Langmead B, Irizarry R, Amdam GV, Feinberg AP: **Reversible switching between epigenetic states in honeybee behavioral subcastes.** *Nat Neurosci* 2012, **15**(10):1371-1373.
- Mazigo HD, Nuwaha F, Kinung'hi SM, Morona D, Pinot de Moira A, Wilson S, Heukelbach J, Dunne DW: **Epidemiology and control of human schistosomiasis in Tanzania.** *Parasit Vectors* 2012, **5**:274.
- Walker AJ: **Insights into the functional biology of schistosomes.** *Parasit Vectors* 2011, **4**:203.
- Blouin M, Thuillier V, Cooper B, Amarasinghe V, Cluzel L, Araki H, Grunau C: **No evidence for large differences in genomic methylation between wild and hatchery steelhead (*Oncorhynchus mykiss*).** *Can J Fish Aquat Sci* 2010, **67**:217-224.
- Raddatz G, Guzzardo PM, Olova N, Fantappie MR, Rampp M, Schaefer M, Reik W, Hannon GJ, Lyko F: **Dnmt2-dependent methylomes lack defined DNA methylation patterns.** *Proc Natl Acad Sci USA* 2013, **110**(21):8627-8631.
- Fantappie MR, Gimba ER, Rumjanek FD: **Lack of DNA methylation in *Schistosoma mansoni*.** *Exp Parasitol* 2001, **98**(3):162-166.
- Geyer KK, Rodriguez Lopez CM, Chalmers IW, Munshi SE, Truscott M, Heald J, Wilkinson MJ, Hoffmann KF: **Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*.** *Nat Commun* 2011, **2**:242.
- Deleury E, Dubreuil G, Elangovan N, Wajnberg E, Reichhart JM, Gourbal B, Duval D, Baron OL, Gouzy J, Coustau C: **Specific versus non-specific immune responses in an invertebrate species evidenced by a comparative de novo sequencing study.** *PLoS One* 2012, **7**(3):e32512.
- Goecks J, Nekrutenko A, Taylor J: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**(8):R86.

29. Raghavan N, Tettelin H, Miller A, Hostetler J, Tallon L, Knight M: **Nimbus (Bgl): an active non-LTR retrotransposon of the *Schistosoma mansoni* snail host *Biomphalaria glabrata*.** *Int J Parasitol* 2007, **37**(12):1307–1318.
30. Ittiprasert W, Knight M: **Reversing the resistance phenotype of the *Biomphalaria glabrata* snail host *Schistosoma mansoni* infection by temperature modulation.** *PLoS Pathog* 2012, **8**(4):e1002677.
31. Grunau C, Clark SJ, Rosenthal A: **Bisulfite genomic sequencing: systematic investigation of critical experimental parameters.** *Nucleic Acids Res* 2001, **29**(13):E65–65.
32. Boyd VL, Zon G: **Bisulfite conversion of genomic DNA for methylation analysis: protocol simplification with higher recovery applicable to limited samples and increased throughput.** *Anal Biochem* 2004, **326**(2):278–280.
33. Li LC, Dahiya R: **MethPrimer: designing primers for methylation PCRs.** *Bioinformatics* 2002, **18**(11):1427–1431.
34. Eads CA, Laird PW: **Combined bisulfite restriction analysis (COBRA).** *Methods Mol Biol* 2002, **200**:71–85.
35. Grunau C, Schattevoy R, Mache N, Rosenthal A: **MethTools—a toolbox to visualize and analyze DNA methylation data.** *Nucleic Acids Res* 2000, **28**(5):1053–1058.
36. Glastad KM, Hunt BG, Yi SV, Goodisman MA: **DNA methylation in insects: on the brink of the epigenomic era.** *Insect Mol Biol* 2011, **20**(5):553–565.
37. Robinson KL, Tohidi-Esfahani D, Lo N, Simpson SJ, Sword GA: **Evidence for widespread genomic methylation in the migratory locust, *Locusta migratoria* (Orthoptera: Acrididae).** *PLoS One* 2011, **6**(12):e28167.
38. Fraley C, Raftery AE: **Model-based clustering, discriminant analysis, and density estimation.** *J Am Statist Association* 2002, **97**:611–631.
39. Fraley C, Raftery AE, Scrucca L: **mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.** *Technical Report No 597.* Seattle: Department of Statistics, University of Washington; 2012.
40. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674–3676.
41. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH: **Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2.** *Science* 2006, **311**(5759):395–398.
42. Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science* 2009, **324**(5929):929–930.
43. Dahl C, Gronbaek K, Guldberg P: **Advances in DNA methylation: 5-hydroxymethylcytosine revisited.** *Clin Chim Acta* 2011, **412**(11–12):831–836.
44. Hendrich B, Tweedie S: **The methyl-CpG binding domain and the evolving role of DNA methylation in animals.** *Trends Genet* 2003, **19**(5):269–277.
45. Wang S, Bao Z, Hu X, Shao M, Zhang L, Hu J: **Two novel elements (CFG1 and PYG1) of Mag lineage of Ty3/Gypsy retrotransposons from Zhikong scallop (*Chlamys farreri*) and Japanese scallop (*Patinopecten yessoensis*).** *Genetica* 2008, **133**(1):37–46.
46. Lechner M, Marz M, Ihling C, Sinz A, Stadler PF, Krauss V: **The correlation of genome size and DNA methylation rate in metazoans.** *Theory Biosci* 2013, **132**(1):47–60.
47. Falckenhayn C, Boerjan B, Raddatz G, Frohme M, Schoofs L, Lyko F: **Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*.** *J Exp Biol* 2012, **216**(8):1423–1429.
48. Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y, et al: **Single base-resolution methylome of the silkworm reveals a sparse epigenomic map.** *Nat Biotechnol* 2010, **28**(5):516–520.
49. Jais KW, Bachman KE, Suzuki H, Ting AH, Rhee I, Yen RW, Baylin SB, Schuebel KE: **De novo CpG island methylation in human cancer cells.** *Cancer Res* 2006, **66**(2):682–692.
50. Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE: **Functional CpG methylation system in a social insect.** *Science* 2006, **314**(5799):645–647.
51. Mohan KN, Chandra HS: **Isolation and analysis of sequences showing sex-specific cytosine methylation in the mealybug *Planococcus lilacinus*.** *Mol Genet Genomics* 2005, **274**(6):557–568.
52. Phalke S, Nickel O, Walluscheck D, Hortic F, Onorati MC, Reuter G: **Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2.** *Nat Genet* 2009, **41**(6):696–702.
53. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W: **Molecular basis of base substitution hotspots in *Escherichia coli*.** *Nature* 1978, **274**(5673):775–780.
54. Bird AP: **DNA methylation and the frequency of CpG in animal DNA.** *Nucleic Acids Res* 1980, **8**(7):1499–1504.
55. Glastad KM, Hunt BG, Goodisman MA: **Evidence of a conserved functional role for DNA methylation in termites.** *Insect Mol Biol* 2013, **22**(2):143–154.
56. Elango N, Hunt BG, Goodisman MA, Yi SV: **DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*.** *Proc Natl Acad Sci USA* 2009, **106**(27):11206–11211.

doi:10.1186/1756-3305-6-167

Cite this article as: Fneich et al.: 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*. *Parasites & Vectors* 2013 **6**:167.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*

Sara Fneich, Nolwenn Dheilly, Coen Adema, Anne Rognon, Michael Reichelt, Jan Bulla, Christoph Grunau and Céline Cosseau

Les données supplémentaires :

Additional file 1: Blast analysis on methylation machinery candidates

DNMT1 tblastN results (query sequence: AAI26228, 1632AA):

Localisation on query sequence	Localisation on bg contig sequence	% positives	Conserved domains
0-362		<=40%	pfam06464, DMAP1-binding Domain
362-426	Contig19819.1 2302-2018	42%	
423-467	Contig19819.1 1666-1532	73%	pfam12047 Cytosine specific DNA
467-550	Contig15589.2 4796-4545	80%	methyltransferase replication foci domain
531-676	Contig15589.2 4227-3727	51%	
669-706	Contig15589.2 3189-3076	86%	pfam02008, CXXC zinc finger domain
706-769	Contig15589.2 2082-1876	49%	
776-850	Contig15589.2 881-660	69%	cd04760, Bromo domain, present in DNA
849-921	Contig15589.1 26165-25947	59%	(Cytosine-5)-methyltransferases
929-978	Contig15589.1 25431-25279	60%	cd04711, Bromo domain, present in DNA
980-1039	Contig15589.1 24988-24812	76%	(Cytosine-5)-methyltransferases
1039-1120	Contig15589.1 24549-24304	69%	
1152-1235	Contig15589.1 23501-23252	72%	cl16911, S-adenosylmethionine-dependent methyltransferases
1231-1316	Contig15589.1 22740-22483	96%	
1309-1370	Contig15589.1 21975-21787	87%	
1366-1406	Contig15589.1 21344-21222	80%	
1405-1471	Contig15589.1 20414-20214	88%	

1473-1552	Contig15589.1	87%	
1553-1592	20118-19882 Contig15589.1 19659-19540	97%	
1591-1613	Contig15589.1 19309-19241	60%	

DNMT2 tblastN results (query sequence: CAG29312, 391AA):

Localizations on query sequence	Localisation on bg contig sequence	% positives	Conserved domains
4-47	Contig6274.3 9738-9610	68%	cd00315, Cytosine-C5 specific DNA methylases cl16911S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)
22-60	Contig6274.3 7385-7269	69%	
49-106	Contig6274.3 6882-6709	67%	
108-168	Contig6274.3 5192-5010	75%	
168-253	No blast results		
253-295	Contig6274.3 3452-3321	72%	cl16911, S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase)
311-358	Contig6274.3 1711-1568	60%	
359-389	Contig6274.3 881-789	80%	

DNMT3 tblastN results (query sequence: NP_787046.1, 845AA): No significant results

Tet1 tblastN results (query sequence: NP_085128.2, 2136AA):

Localizations on query sequence	Localisation on bg contig sequence	% positives	Conserved domains
0-1488	No blast results		cl03401, CXXC zinc finger domain
1488-1601	Contig18827.1 4390-3917	51%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily
1604-1639	Contig18827.1 2928-2722	72%	
1638-1685	Contig18827.1 1356-1213	79%	
1456-1488	Contig21.52 41322-41224	81%	
1488-1557	Contig21.52 40969-40763	67%	pfam12851, Oxygenase domain of the

1559-1601	Contig21.52 40623-40495	81%	2OGFeDO superfamily
1604-1639	Contig21.52 38529-38422	81%	
1638-1685	Contig21.52 37121-36978	79%	
1700-1738	Contig21.52 35681-35565	81%	
1738-1996	No blast results		
1996-2062	Contig21.52 33830-33630	76%	

Tet2 tblastN results (query sequence: NP_001120680.1, 2002AA)

Localizations on query sequence	Localisation on bg contig sequence	% positives	Conserved domains*
0-1250	No blast results		
1199-1311	Contig18827.1 4390-3917	55%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily
1343-1395	Contig18827.1 1371- 1213	75%	
1319-1349	Contig18827.1 2814- 2722	70%	
1096-1167	43136-42942	76%	
1167-1199	Contig21.52 41322-41224	84%	
1199-1268	Contig21.52 40969-40760	71%	
1269-1311	Contig21.52 40623-40495	86%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily
1319-1349	Contig21.52 38514-38422	70%	
1343-1395	Contig21.52 37136-36978	75%	
1409-1446	Contig21.52 35684-35511	68%	
1849-1914	Contig21.52 33830-33633	78%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily

Tet3 tblastN results (query sequence: NP_659430.1, 1660AA)

Localizations on query sequence	Localisation on bg contig sequence	% positives	Conserved domains*
1-759	No blast results		
759-886	Contig18827.1 4390-3887	52%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily
1343-1395	Contig18827.1 1371-1213	75%	
1319-1349	Contig18827.1 2814-2722	70%	
1096-1167	43136-42942	76%	
1167-1199	Contig21.52 41322-41224	84%	
1199-1268	Contig21.52 40969-40760	71%	
1269-1311	Contig21.52 40623-40495	86%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily
1319-1349	Contig21.52 38514-38422	70%	
1343-1395	Contig21.52 37136-36978	75%	
1409-1446	Contig21.52 35684-35511	68%	
1849-1914	Contig21.52 33830-33633	78%	pfam12851, Oxygenase domain of the 2OGFeDO superfamily

*Conserved domains were detected by doing a specialized blast (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) using as query protein the homo sapiens characterized proteins.

MBD1 tblastN results (query sequence: NP_001191066.1): no significant results

MBD2 tblastN results (query sequence: AAC68871.1 411AA):

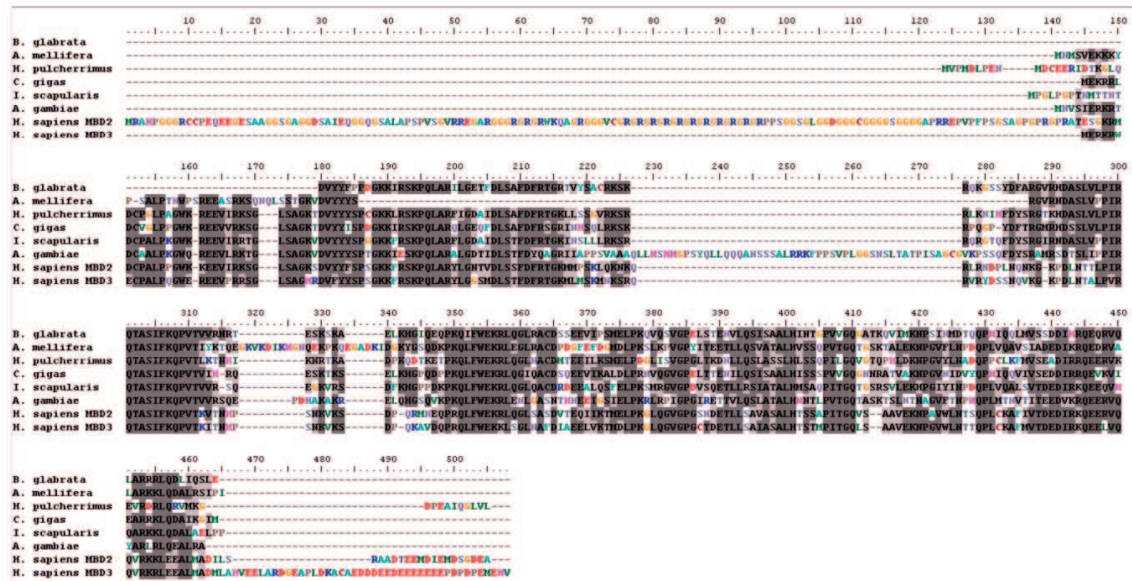
Localizations on query sequence	Localisation on bg contig sequence	% positives	Conserved domains*
0-154			
154-181	Contig7228 9946-9863	100%	cd01396, MeCP2, MBD1, MBD2, MBD3, and MBD4 are members of a protein family that share the methyl-CpG-binding domain (MBD)
179-224	Contig7228 7409-7272	78%	
204-305	Contig7228 7155-6856	53%	
281-310	Contig7228 6226-6137	83%	pfam14048, C-terminal domain of methyl-CpG binding protein 2 and 3
311-357	Contig7228 5753-5607	65%	
357-411			

MBD3 tblastN results (query sequence: AAC68876.1, 291AA):

Localizations on query sequence	Localisation on bg contig sequence	% positives	Conserved domains*
10-37	Contig7228 9946-9863	85%	cd01396, MeCP2, MBD1, MBD2, MBD3, and MBD4 are members of a protein family that share the methyl-CpG-binding domain (MBD)
35-92	Contig7228 7409-7239	67%	
94-137	Contig7228 7050-6916	71%	pfam14048, C-terminal domain of methyl-CpG binding protein 2 and 3
137-166	Contig7228 6226-6137	80%	
167-213	Contig7228 5753-5607	63%	

MBD4 tblastN results (query sequence: AAC68879.1, 580AA): only the carboxy terminal part of the query protein matches with a target in the *B. glabrata* genome. This carboxy terminal part of AAC68879.1 contains a domain involved in endonuclease activity (cl14786). The coverage is too low to conclude for the presence of a MBD4 in *B. glabrata* genome.

Additional file 2: Sequence Alignments of MBD2/3 protein families.



Additional file 3: Classification of GO functional categories in housekeeping, inducible or non informative subcategories.

Bg Gua cDNA (≥ 500 pb) were divided into 11 cDNA subsets based on their CpGo/e ratio. Each subset was tested for enrichment in functional category using the Blast2GO software. Over-represented or underrepresented GO terms (p value < 0.05) were classified into 3 subcategories: housekeeping GO (underlined in red), non informative GO (not underlined), inducible GO (underlined in yellow).

Under-represented GO terms :

0 < CpG/GpC <= 0,1	0,1 < CpG/GpC <= 0,2	0,2 < CpG/GpC <= 0,3	0,3 < CpG/GpC <= 0,4	0,4 < CpG/GpC <= 0,5	0,5 < CpG/GpC <= 0,6
GTPase activity	chitin metabolic process	carbohydrate binding		protein binding	
GTP binding	serine-type endopeptidase inhibitor activity	carbohydrate metabolic process		protein complex	
extracellular region	chitin binding	serine-type endopeptidase inhibitor activity		intracellular organelle part	
GTP catabolic process	extracellular region part			cytoplasm	
translational elongation	RNA-directed DNA polymerase activity			intracellular protein transport	
ribosome	iron ion binding			cellular component organization or biogenesis at cellular level	
structural constituent of ribosome				cellular macromolecule metabolic process	
carbohydrate binding				hydrolase activity, acting on acid anhydrides	
aminoglycan metabolic process					
serine-type endopeptidase inhibitor activity					

0,6 < CpG/GpC <= 0,7	0,7 < CpG/GpC <= 0,8	0,8 < CpG/GpC <= 0,9	0,9 < CpG/GpC <= 1	> 1
protein complex	mitochondrion	cytoplasmic part		membrane-bounded organelle
nucleolus organelle	intracellular non-membrane-bounded organelle	ATP binding		intracellular membrane-bounded organelle
membrane	protein complex	RNA metabolic process		cell
RNA processing	regulation of primary metabolic process	nuclear part		cell part
intracellular protein transport	nucleoplasm	regulation of primary metabolic process		membrane part
ATP binding	regulation of cellular metabolic process			intrinsic to membrane
cytosol	organelle organization			intracellular
endomembrane system	regulation of macromolecule metabolic process			integral to membrane
nucleoplasm	ATP binding			intracellular part
intracellular signal transduction	organelle envelope			intracellular organelle
positive regulation of catalytic activity	transcription, DNA-dependent			nucleus
ATPase activity, coupled	establishment of localization in cell			organelle
carboxylic acid metabolic process	cellular response to stress			transport
RNA binding	macromolecule modification			
nucleocytoplasmic transport				
cell cycle				
phosphorylation				
organelle envelope				
kinase activity				
mitochondrial part				
mRNA metabolic process				
chromatin modification				
regulation of protein modification process				
response to DNA damage stimulus				
neurological system process				
cellular catabolic process				
regulation of transcription, DNA-dependent				
microtubule cytoskeleton				

Over-represented GO terms:

0 < CpG/GpC <= 0,1	0,1 < CpG/GpC <= 0,2	0,2 < CpG/GpC <= 0,3	0,3 < CpG/GpC <= 0,4	0,4 < CpG/GpC <= 0,5
cellular amino acid metabolic process	Arp2/3 protein complex	ATP binding	cellular component organization at cellular level	cellular macromolecule biosynthetic process
cellular response to lipid	ATP binding	ATP hydrolysis coupled proton transport	cellular protein localization	cytosol
fatty acid metabolic process	ATPase binding	cellular component biogenesis	heterogeneous nuclear ribonucleoprotein complex	intracellular signal transduction
intracellular membrane-bounded organelle	cellular protein modification process	Golgi vesicle transport	mitochondrion	mitochondrion nuclear mRNA splicing, via spliceosome
intracellular transport	ligase activity	mitochondrion	nucleolus	nucleoplasm part
lyase activity	peptidylglycine monooxygenase activity	mitotic prometaphase	nucleoside-triphosphatase activity	positive regulation of proteasomal ubiquitin-dependent protein catabolic process
protein complex	protein lipidation	NAD metabolic process	primary cilium	protein complex regulation of gene expression
protein N-linked glycosylation	protein transport	nuclear outer membrane	regulation of RNA metabolic process	RNA binding
		nuclear pore	synapse system development	
		nucleoplasm positive regulation of GTPase activity		
		proteolysis involved in cellular protein catabolic process		
		Rab GTPase activator activity		
		regulation of cytoskeleton organization		
		regulation of glucose transport		
		regulation of macromolecule metabolic process		
		regulation of Ras GTPase activity		
		RNA processing		
		vacuole		
		vitamin A metabolic process		
		endoplasmic reticulum membrane		

0,5 < CpG/GpC <= 0,6	0,6 < CpG/GpC <= 0,7	0,7 < CpG/GpC <= 0,8	0,8 < CpG/GpC <= 0,9	0,9 < CpG/GpC <= 1	> 1
aminoglycan metabolic process	extracellular region	amylase activity	extracellular region	primary amine oxidase activity	ribosome
hydrolase activity, acting on glycosyl bonds	ferroxidase activity	chitin binding chitin metabolic process	carbohydrate binding	carbohydrate binding	structural constituent of ribosome
oxygen binding	polysaccharide binding			GTP binding	translation
oxygen transport oxygen transporter activity	serine-type endopeptidase inhibitor activity tumor necrosis factor receptor binding	extracellular space		quinone binding	structural molecule activity
polysaccharide binding		GTP binding metalloendopep tidase activity			ribonucleoprote in complex
		proteolysis			cytosolic part endocrine pancreas development endocrine system development
		serine-type endopeptidase activity serine-type endopeptidase inhibitor activity sulfotransferase activity tumor necrosis factor receptor binding			GTP binding translational termination
					guanyl ribonucleotide binding guanyl nucleotide binding pancreas development viral genome expression viral transcription translation elongation factor activity
					metalloendopep tidase activity striated muscle thin filament protein metabolic process
					troponin complex protein complex disassembly viral infectious cycle
					calcium ion binding cellular process involved in reproduction

III.4. Résultats complémentaires non publiés

La méthylation de l'ADN, un des porteurs de l'information épigénétique, existe chez l'hôte intermédiaire *B. glabrata*. Ceci suggère que des mécanismes épigénétiques contrôlent l'expression de certains phénotypes chez le mollusque.

La première question biologique que nous nous posons :

Quel rôle joue la méthylation de l'ADN chez le mollusque ?

Dans ce contexte d'interaction hôte / parasite :

Existe-t-il des variations de la méthylation de l'ADN chez l'hôte avant et après l'infestation par le parasite ? Ou autrement dit, est ce que l'épigénome du mollusque varie en fonction de l'infestation par le parasite ?

Pour ces investigations, il était important de passer à une étude plus élargie, échelle de population de mollusques. Afin de comparer les variations de l'épigénome de plusieurs individus, théoriquement il faut que ces individus soient homogènes au niveau génétique. Il serait donc essentiel de tester la clonalité génétique de ces individus.

Pour cela, nous avons utilisé la technique Methylation Sensitive- Amplified Fragment Lenght Polymorphism (MS-AFLP) (Herrera & Bazaga 2011). En effet, cette technique à double avantage, permet d'étudier à la fois le polymorphisme génétique et le polymorphisme épigénétique. Quatre différentes populations de mollusques ont été utilisées pour cette étude :

- 16 individus *Bg* BRE non infestés
- 18 individus *Bg* GUA non infestés
- 9 individus *Bg* GUAi infestés par *Sm* GH2
- 11 individus *Bg* GUAEni exposés à *Sm* GH2 mais non infestés

Cette partie a fait l'objet de stage Matser II, Hanine Haidar, que j'ai co-encadrée en 2012.

Protocole expérimental :

L'ADN génomique de chaque individu mollusque a été extrait par la technique de phénol/Chloroforme (décrite sur le site <http://methdb.univ-perp.fr/epievo/>).

1. 200 ng d'ADNg ont été digérés par chacune des enzymes isoschizomères *MspI* et *HpaII*, à coupure fréquente, chacune en combinaison avec la même enzyme *EcoRI* à coupure rare, pendant 2 heures à 37°C. Ce qui résulte deux réactions en parallèle pour chaque échantillon: une digestion *MspI/EcoRI* et une autre digestion *HpaII/EcoRI*.
2. Après l'inactivation des enzymes pendant 20 minutes à 60°C, la ligation des adaptateurs a été effectuée pendant 2 heures à 37°C. Ceci a été effectué en utilisant les amorces adaptées aux sites générés par la digestion enzymatique (tableau 3).

	Séquence des amorces	Séquence des adaptateurs
EcoRI	Amorce fluorescente 5'-GACTGCGTACCAATTCCTG-3'	Linker1 5'-CTCGTAGACTGCGTACC-3'
		Linker 2 5'-AATTGGTACGCAGTCTAC-3'
MspI/HpaII	5'-GATGAGTCTAGAACGGTCC-3'	Linker1 5'-GACGATGAGTCTAGAA-3'
		Linker2 5'-CGTTCTAGACTCATC-3'

Tableau 3 : Les séquences des amorces et des adaptateurs utilisés au cours du protocole expérimental de la MS-AFLP.

3. La réaction de PCR est effectuée ensuite, en utilisant une dilution 1/20^{ème} de la réaction du mix digestion / ligation. L'amorce fluorescente est utilisée pour une détection ultérieure par le séquenceur capillaire CEQ800.
4. L'analyse des résultats a été réalisée en deux étapes. Tout d'abord, les produits obtenus après PCR ont été visualisés par une électrophorèse sur gel d'agarose 2%. Ensuite, les échantillons sont passés sur le séquenceur capillaire. Le master mix est

préparé par 1µl de chaque produit PCR avec 40µl SLS et 0.18 µl du marqueur de taille de 400 pb. Le résultat est finalement visualisé par un profil de bandes qui sont par la suite analysées (figure 23).

Au niveau populationnel : Dans le cas où les profils de *MspI* (ou bien de *HpaII*) diffèrent entre les individus, ceci implique que les individus de cette population présentent des différences au niveau génétique. Des outils de génétique de population sont utilisés à ce stade là.

Au niveau individuel : Dans le cas où les profils de *MspI* et *HpaII* sont différents au sein d'un même individu, ceci montre qu'il existe des sites méthylés dans son génome.

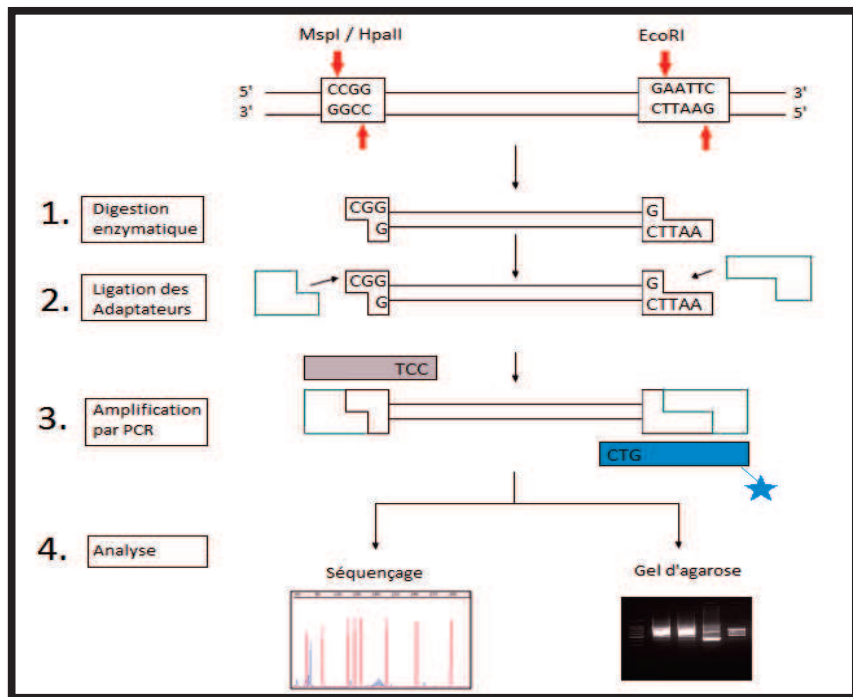


Figure 23 : Les différentes étapes expérimentales de la technique MS-AFLP.

Résultats :

Bien que les mollusques soient élevés au laboratoire depuis trente ans environ, la MS-AFLP a montré que les individus appartenant à chacune des populations de mollusque *Bg* BRE ou *Bg* GUA sont génétiquement diversifiés entre eux. Avec cette

diversité génétique importante dans chacune des populations *Bg* BRE et *Bg* GUA, la diversité épigénétique était difficile à conclure.

Etude de la structure des populations :

La diversité génétique entre les individus d'une même souche (et même population) de mollusque n'était pas négligeable. Le profil de bandes de chaque individu obtenu par MS-AFLP a été analysé.

Nous avons obtenu 146 marqueurs polymorphes pour les 54 individus de *B. glabrata* des 4 différentes populations. L'analyse en composantes principales ACP montre 31.2% de variations sur les deux axes (figure 24). L'axe 1 (18.9%) sépare clairement les deux différentes souches de mollusques *Bg* BRE et *Bg* GUA. Alors que l'axe 2 (12.3%) nous permet de distinguer 3 groupes différents au sein de la même souche *Bg* GUA. La population des *Bg* GUA non infestés (cercles bleus) couvre une aire mixte incluant les autres populations de *Bg* GUA. Cependant, cette population n'atteint pas toute la localisation des *Bg* GUAEni exposés non infestés (losanges jaunes).

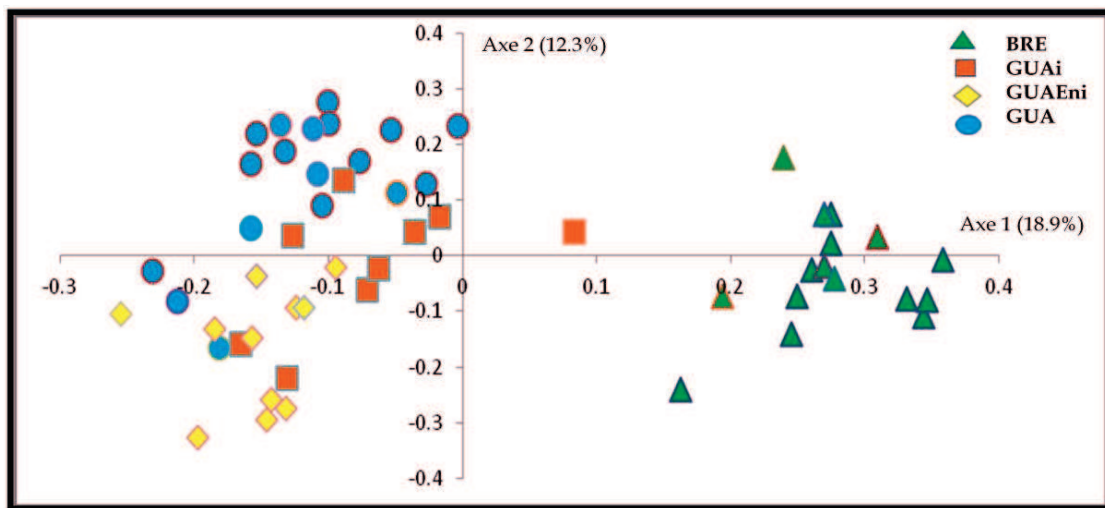


Figure 24 : La distribution des 54 individus de différentes populations de *B. glabrata* obtenus par ACP. Chaque symbole correspond à un individu mollusque appartenant à une population différente : *Bg* BRE non infestés (triangles verts), *Bg* GUAi infestés (rectangles oranges), *Bg* GUAEni exposés non infestés (losanges jaunes) et *Bg* GUA non infestés (cercles bleus). 31.2% de ces variables sont représentés dans les résultats. L'axe des abscisses représente les 18.9% de ces variations et l'axe des ordonnées représente les 12.3%.

Ensuite, la diversité génétique au sein de chacune des populations a été étudiée. Afin d'estimer les diversités nous nous sommes servi de l'indice de Shannon. Et finalement, le test de Wilcoxon a été utilisé pour comparer les 4 populations entre elles (Bonin *et al.* 2007). L'avantage de ce test c'est qu'il prenne en compte le niveau de différence à l'intérieur de chaque population. Les résultats ont montré que la population *Bg* GUAEni exposés non infestés est la plus diversifiée génétiquement par rapport aux autres populations de mollusques.

Chapitre IV

Discussion

Chapitre IV : Discussion

L'objectif principal de cette thèse était de montrer la contribution des mécanismes épigénétiques dans la production des variants phénotypiques dans le modèle d'interaction *B. glabrata* / *S. mansoni*. Les interactions hôte / parasite constituent un bon modèle pour cette étude parce qu'elles se caractérisent par une dynamique coévolutive (Van Valen 1974). Dans une telle interaction, cet équilibre dynamique nécessite la production de variants phénotypiques aussi bien de la part de l'hôte *B. glabrata* que de la part du parasite *S. mansoni* (Frank 2000). Nous proposons l'hypothèse que l'épigénétique soit une source de production de variants phénotypiques dans cette interaction, assurant une adaptation rapide et une coévolution permanente des 2 partenaires.

Les études effectuées au laboratoire ont mis en évidence un polymorphisme de compatibilité dans l'interaction *S. mansoni* / *B. glabrata* (Théron & Coustau 2005). Ce polymorphisme est traduit par des éléments moléculaires clés des deux partenaires, déterminant la compatibilité / incompatibilité d'une interaction (Mitta *et al.* 2012).

La famille multigénique *SmPoMuc* est considérée comme un candidat moléculaire déterminant pour les phénomènes de compatibilité (Roger *et al.* 2008c). En effet, les protéines *SmPoMucs* ont montré une interaction avec des antigènes du mollusque par des approches de co-immunoprécipitation (Moné *et al.* 2010). Appartenant à une famille multigénique, les *SmPoMucs* pourraient être régulées par différents mécanismes à l'origine de leur diversification (Roger *et al.* 2008b). Récemment, les études moléculaires ont suggéré que les modifications des histones au niveau des promoteurs de ces gènes soient responsables du contrôle de l'expression de certains des variants *SmPoMucs* chez *S. mansoni* (Cosseau *et al.* 2009; Cosseau *et al.* 2010; Perrin *et al.* 2013). Ces données constituent l'état de l'art du début de ma thèse en ce qui concerne le parasite et l'implication de l'épigénétique dans sa plasticité phénotypique. Du côté du mollusque *B. glabrata*, aucune modification chimique de l'ADN n'avait été décrite avant le début de ma thèse.

Nos résultats apportent de nouveaux éléments quant à la compréhension des mécanismes d'adaptation dans un système nécessitant une évolution qui est relativement rapide. Les deux chapitres II et III de la thèse sont consacrés aux résultats de l'implication de l'épigénétique dans le succès du parasite et de la mise en évidence de la méthylation de l'ADN chez le mollusque.

Ce présent chapitre de Discussion est consacré surtout aux parties non publiées autour des travaux effectués au cours de la thèse.

IV.1. Epigénétique et adaptation chez *S. mansoni*

Dans le chapitre II de la thèse nous avons effectué des croisements entre les deux différentes souches du parasite, provenant de deux origines géographiques différentes (souches brésiliennes et souches guadeloupéennes). Le croisement a généré des hybrides parasites qui ont montré :

Résultat I : Des nouveaux variants moléculaires de *SmPoMucs* qui sont exprimés. Ces variants phénotypiques sont soumis à un contrôle épigénétique par des modifications des histones autour des promoteurs (Publication N°2). Contrairement à la transmission mendélienne de l'ADN, l'héritabilité des structures chromatinienne pourrait suivre une transmission mendélienne ou même non-mendélienne (Publication N°2).

Résultat II : Une virulence ou hétérosis observée sur l'hôte définitif (hamsters) et une augmentation de la compatibilité avec l'hôte intermédiaire *B. glabrata*.

Concernant le résultat I :

Vers une étude fonctionnelle des *SmPoMucs*

Les *SmPoMucs* sont des glycoprotéines hautement diversifiées, qui montrent une expression différentielle entre les différentes souches du parasite compatibles et incompatibles (Roger *et al.* 2008c). Les glycoprotéines se retrouvent d'une façon générale à la surface des cellules chez les parasites, jouant un éventuel rôle dans la protection du parasite par échappement aux mécanismes de défense de l'hôte (Theodoropoulos *et al.* 2001; Verstrepen & Fink 2009). Ces *SmPoMucs* sont

particulièrement intéressantes car le maintien de la dynamique coévolutive dans les systèmes hôte / parasite reposerait en partie sur la diversification et / ou polymorphisme des molécules jouant des rôles clés au cœur de l'interaction (Van Valen 1974). Le polymorphisme et le statut de glycosylation sont deux caractéristiques importantes pour la reconnaissance des antigènes de l'hôte.

Des approches de co-immunoprécipitation, ont montré que plusieurs glycoprotéines de *S. mansoni*, dont les *SmPoMucs* interagissent avec des récepteurs immunitaires hautement diversifiées du mollusque (Moné *et al.* 2010). Les récepteurs immunitaires co-précipités du mollusque sont principalement des lectines, dont les FREPs. Les lectines se caractérisent par leur capacité à lier des motifs glucidiques d'autres molécules. Ces récepteurs sont capables de reconnaître des motifs moléculaires caractéristiques des pathogènes en remplissant une fonction PRR (Pattern Recognition Receptors). Cette reconnaissance déclencherait une réaction immunitaire innée chez l'hôte (Dam & Brewer 2010).

Evidemment, l'interaction *B. glabrata* / *S. mansoni* ne se reposerait pas uniquement sur la reconnaissance FREPs / *SmPoMucs* mais aussi d'autres facteurs, cette reconnaissance n'est qu'une partie du puzzle. Moné et ses collègues ont montré des réactions oxydante de l'hôte / antioxydante du parasite qui sembleraient être au cœur de l'interaction aussi. En effet, ils ont mis en évidence des effecteurs responsables de la destruction du parasite, les ROS ou espèces réactives de l'oxygène chez *B. glabrata*. Également, *S. mansoni* possède des systèmes antioxydants, ROS *scavengers* qui permettent de contrecarrer le système immunitaire de l'hôte (Moné *et al.* 2011).

Ayant conscience que l'équilibre de la reine rouge rencontrée dans un système hôte / parasite est multifactorielle, nous nous sommes focalisés au cours de la thèse uniquement sur les *SmPoMucs*. Les recherches menées jusqu'à présent, étaient consacrées surtout à comprendre les mécanismes de diversification intrigante de ces molécules. Cependant, la fonction moléculaire n'est pas encore claire.

A l'aide d'un anticorps spécifique reconnaissant les *SmPoMucs*, nous avons pu étudier les différents profils d'expression dans différentes souches du parasite ainsi que dans différentes conditions.

Le croisement des deux souches *Sm* BRE et *Sm* GH2 a montré que les hybrides de toutes les générations expriment de nouveaux variants de *SmPoMuc*, ce qui se traduit par des bandes supplémentaires détectées en Western blot. Egalement, le traitement des miracidia avec un agent épimutagène, la Trichostatin A (TSA), a montré l'apparition de nouveaux variants de *SmPoMucs* détectées par la même technique (gardant toujours les mêmes conditions). Dans les deux conditions, l'hybridation des deux souches de parasites et le traitement par un agent épimutagène, a conduit à observer une meilleure compatibilité des parasites vis-à-vis du mollusque (Publication N°2).

Ces observations nous ont permis de suggérer en premier lieu, que l'augmentation du nombre de variants des *SmPoMucs* augmenterait la compatibilité du parasite. Afin de tester cette hypothèse, nous avons comparé le profil d'expression des *SmPoMucs* des deux souches *Sm* BRE et *Sm* GH2 avec la souche *Sm* LE. En effet, *Sm* LE est la souche de *S. mansoni* qui a la capacité la plus élevée d'infestation, avec des prévalences de ~100% sur 4 différentes souches de mollusques (figure 25) (Theron *et al.* 2014).

Le résultat du Western blot a montré un profil d'expression de *SmPoMucs* chez *Sm* LE, qui est plus réduit en bandes que celui des autres souches testées (figure 26). Cette approche ne fournit pas d'informations sur les séquences exactes en acides aminés et les modifications post-traductionnelles des protéines détectées. Cependant, elle montre clairement la présence de variants moléculaires différents entre les différentes souches de parasite. En ce sens là, il pourrait être intéressant d'envisager des analyses plus poussées sur ces différents variants.

Nos résultats obtenus jusqu'à présent nous ont amenés à deux contradictions :

(i) Au vue de nos résultats, il semble qu'une augmentation du nombre de variants de *SmPoMucs* augmenterait la compatibilité du parasite vis-à-vis du mollusque, mais ce n'était pas le cas pour la souche *Sm LE*.

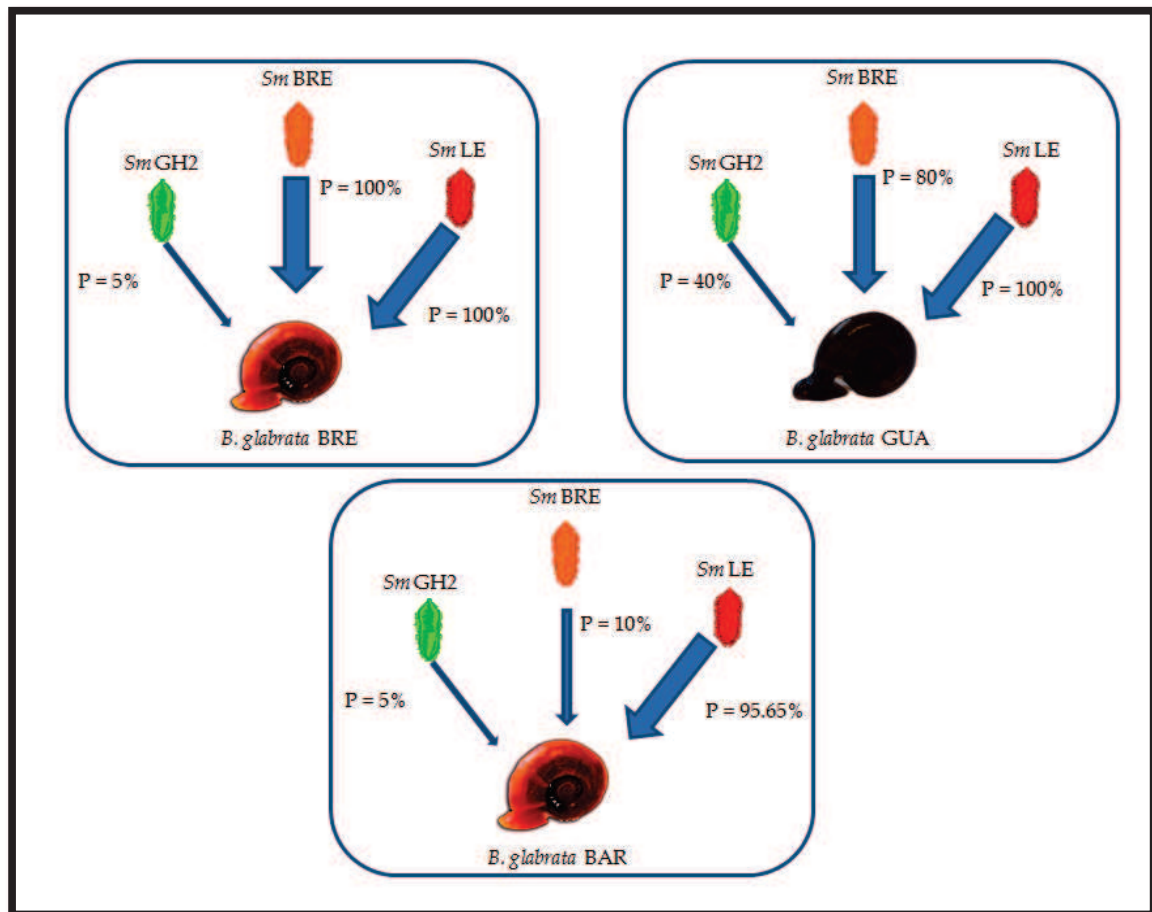


Figure 25 : Représentation schématique des compatibilités entre les différentes souches de *S. mansoni*. Les prévalences obtenues suite à l'infestation de chacune des souches de *B. glabrata* avec chacune des souches de *S. mansoni* (chaque mollusque a été exposé à 20 miracidia de chaque souche / 30 à 50 mollusques ont été utilisés pour chaque condition).

(ii) Revenant à notre marqueur de l'étude *SmPoMuc* 3.1 (r1r2) : les études antérieures effectuées au laboratoire ont montré que ce gène n'est exprimé que chez les souches incompatibles avec le mollusque. Ceci a permis de suggérer que ce variant *SmPoMuc*

3.1 (r1r2) est facilement reconnu par l'hôte et c'est la raison pour laquelle les souches compatibles ne l'exprimeraient pas. En revanche, nous avons remarqué (Publication N°2) que les hybrides, qui ont montré de fortes compatibilités avec le mollusque, expriment aussi le variant *SmPoMuc* 3.1 (r1r2) qui est supposé être reconnu par le mollusque.

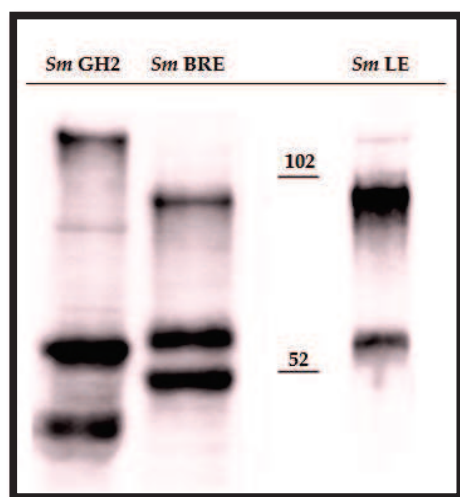


Figure 26 : Profil d'expression de *SmPoMucs* de différentes souches de *S. mansoni*. Western blot effectué sur les trois souches de parasite *Sm* GH2, *Sm* BRE et *Sm* LE à l'aide de l'anti-*SmPoMucs*. Extraits protéiques de 1000 miracidia par puit.

Tous ces résultats contradictoires vont dans le sens que les *SmPoMuc* auraient un rôle qui ne soit pas restreint aux étapes de reconnaissance du parasite par le mollusque. Je souhaite toutefois proposer deux hypothèses quant aux rôles possibles pour cette famille multigénique.

(i) L'augmentation du nombre de variants *SmPoMucs* observée après le croisement ou bien le traitement à la TSA (publication N°2), permet de proposer l'hypothèse que le parasite, afin de réussir à infester dans de telles conditions, augmenterait la diversité moléculaire des variants de *SmPoMucs*. Dans ces conditions, nous pouvons imaginer une stratégie de « dilution ». En effet, après la pénétration du miracidium dans le pied du mollusque, un nombre élevé des variants de *SmPoMucs* seraient exprimés autour du parasite. La concentration des variants serait suffisamment élevée pour empêcher la reconnaissance anticorps du mollusque / le(s) variant(s) habituellement reconnu(s).

(ii) Les *SmPoMucs* sont des protéines hautement glycosylées, possédant des unités répétées VNTR qui semblent être favorables aux O-glycosylation (Roger *et al.* 2008a). Ayant les caractéristiques des mucines, les *SmPoMucs* pourraient assurer un rôle de pénétration du parasite dans le mollusque (Theodoropoulos *et al.* 2001). Sachant que les mucines sont des constituants principaux du mucus, nous supposons qu'elles pourraient faciliter le déplacement du parasite dans les tissus du mollusque.

Ces hypothèses restent difficiles à être testées expérimentalement. Afin d'initier une étude fonctionnelle de ces gènes, nous avons procédé de la technique classique interférence ARN ou RNAi. La technologie du RNAi a été développée chez *S. mansoni* il y a 10 ans environ (Skelly *et al.* 2003). Ce projet a été initié en collaborant avec David Duval (MCF au laboratoire).

Des siRNAs ciblant la région conservée de tous les groupes de *SmPoMucs* ont été synthétisés. Cette technique permettrait une interférence des siRNAs avec les ARN messagers cible, ce qui va conduire à la dégradation des ARNm ou bien diminuer la traduction en protéines. Le but de cette étude est de voir quelle fonction pourrait avoir les *SmPoMucs* dans le contexte de la compatibilité du parasite vis-à-vis du mollusque. En premier lieu, nous avons mis au point l'application de la technique sur les œufs du parasite. Des études phénotypiques seront envisagés par la suite, afin d'évaluer l'éventuel effet de la diminution des *SmPoMucs* sur la compatibilité.

Brièvement, les résultats à l'heure actuelle sont encore préliminaires. Nous avons observé une diminution de la transcription des *SmPoMucs* par RT-qPCR entre le Jour 1 et le Jour 4 après traitement. Ces travaux sont actuellement poursuivis.

Concernant le résultat II :

La virulence sur l'hôte définitif (hamsters)

La virulence pourrait être expliquée par de nombreuses définitions. Chaque définition dépend du thème et du modèle de recherche. Dans notre étude, la virulence de *S. mansoni*, est traduite par les effets néfastes que le parasite a provoqués sur la survie de ses hôtes. Deux hypothèses ont été émises pour expliquer l'évolution du parasite par virulence.

(i) Des études considèrent que la virulence est une « maladaptation » du parasite, liée à une association hôte / parasite qui est récente. En effet, la mortalité de l'hôte diminue la durée de l'infection, et ceci contrarie le profit du parasite. D'après cette hypothèse, l'évolution du parasite devrait mener à une réduction de la virulence. Cette réduction aurait un effet bénéfique pour l'hôte et pour le parasite (Thomas *et al.* 2010).

(ii) En revanche, d'autres études montrent que la virulence du parasite est une « adaptation ». Bien que la virulence réduise la durée de l'infection de l'hôte, le parasite reste bénéficiaire. La virulence pourrait être considérée comme un trait pléiotrope, qui affecte d'autres composantes du phénotype du parasite (Bull 1994). Parmi les phénotypes associés à la virulence, l'augmentation de la transmission du parasite et de la fécondité. D'après cette hypothèse, la virulence ne provoque pas la mortalité directe des hôtes, par contre elle provoque des infections à longues durée (Anderson & May 1982).

Nos résultats obtenus après le croisement des deux souches de parasites ont montré que les hybrides ont provoqué la mortalité des hamsters. Ainsi, nous avons observé une mortalité remarquable qui a augmenté jusqu'à atteindre les 50% des hamsters infestés par les parasites de la génération F2. Ensuite, le taux de mortalité des hamsters diminue à la génération F3. Ceci s'explique parfaitement par la première hypothèse où le parasite se sert du phénotype « virulence », pour l'association récente hybrides (*Sm* BRE x *Sm* GH2) / hôtes. Afin de mieux s'adapter à son hôte, les parasites des générations suivantes réduisent leur virulence.

Cette virulence est traduite par l'augmentation des prévalences et des intensités au niveau des hôtes intermédiaires et par la mortalité au niveau des hôtes définitifs. Ainsi, elle est accompagnée par des variants moléculaires exprimées et qui seraient essentielles pour cette interaction (Publication N°2).

Cette observation de virulence correspond au phénomène hétérosis ou vigueur hybride.

Hétérosis ou vigueur hybride

L'hybride est le produit du croisement des individus appartenant à deux variétés, races, ou espèces différentes. Étant donné que l'hybridation est un mélange de deux génomes distincts, elle pourrait avoir un impact important sur la diversification des espèces et l'acquisition de nouveaux phénotypes (Huyse *et al.* 2009).

Les phénomènes d'hétérosis ou de vigueur hybride sont l'une des conséquences de l'hybridation. Elle désigne le gain des performances phénotypiques des hybrides par rapport à leurs parents en termes de biomasse, fertilité, résistance et adaptation (Shull & Gowen 1952; Eggan *et al.* 2001; Hochholdinger & Hoecker 2007; Baranwal *et al.* 2012). Dans la littérature, un bon nombre d'exemples sont cités. Par exemple, le croisement des deux espèces de grenouille *Rana esculenta* x *R. lessonae*, produit des têtards qui sont plus résistants aux parasites (Planade *et al.* 2009). Ce phénomène est largement exploité en agriculture afin d'améliorer de nombreuses variétés de céréales et des plantes potagères (Swanson-Wagner *et al.* 2006). Le croisement de différentes variétés de riz et de maïs, est fortement utilisé dans le monde entier. Ils estiment une augmentation de 15 à 20% de la production par an (figure 27) (Nuruzzaman *et al.* 2002; Swanson-Wagner *et al.* 2006).

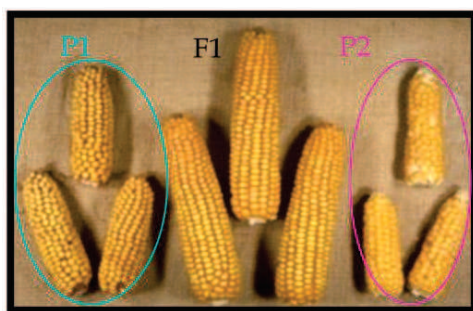


Figure 27 : L'effet de vigueur hybride ou hétérosis. Les épis de maïs de la F1 sont plus longs, murissent plus rapidement et produisent plus de graines que les épis parentaux, d'après (Swanson-Wagner *et al.* 2006).

L'effet hétérosis a commencé à être étudié depuis plus d'un siècle. Il a été décrit pour la première fois par Charles Darwin en 1876 après ses observations de la progéniture du maïs croisé (Johnston 1992). Ce phénomène a été redécouvert par

Shull en 1908 qui a inventé le terme hétérosis pour la première fois en 1914 (Shull 1914; Shull 1948; Shull & Gowen 1952).

Bien que ce phénomène soit largement utilisé et appliqué (Sorensen *et al.* 2008; Planade *et al.* 2009), il a été toujours considéré comme un phénomène « particulier ». En effet, les mécanismes moléculaires expliquant ce phénomène biologique ne sont pas clairs jusqu'à présent.

Les résultats de la vigueur hybride ont été classiquement expliqués par des hypothèses reposant sur des mécanismes génétiques depuis la découverte jusqu'à nos jours. Trois hypothèses sont couramment invoquées :

(i) La dominance : cette hypothèse explique l'hétérosis par l'action de complémentation des allèles dominants supérieurs des deux parents homozygotes sur les allèles récessifs défavorables (délétères) au niveau de différents loci. (ii) La superdominance : cette hypothèse suggère la supériorité d'un trait à l'état hétérozygote par comparaison à l'état homozygote des deux parents (Hochholdinger & Hoecker 2007). (iii) L'épistasie : cette dernière considère les interactions entre les différents gènes au niveau de deux ou plusieurs loci (Bagheri & Jelodar 2010).

Il est important de garder à l'esprit que ces hypothèses de génétique qualitative, ne sont pas directement associées aux traits phénotypiques quantitatives et moléculaires. Avec les nouvelles avancées de la biologie moléculaire, les études transcriptomiques ont permis d'expliquer certains aspects d'hétérosis par différents modèles d'expression de gènes qui sont corrélés à la vigueur des hybrides (Baranwal *et al.* 2012).

Plus récemment, l'éventuelle implication des mécanismes épigénétiques est mise en évidence dans le phénomène d'hétérosis (Han *et al.* 2008; Baranwal *et al.* 2012). Il a été démontré chez certaines plantes, telles que le coton *Gossypium hirsutum* L., l'arabette des dames *Arabidopsis thaliana* et d'autres, que le taux de méthylation des cytosines change dans les lignées hybrides par rapport aux lignées parentales (Zhao *et al.* 2008; Shen *et al.* 2012).

Chez le coton, ils ont comparé le taux de méthylation entre les hybrides non hétérotiques et des hybrides hétérotiques par MS-AFLP. Leurs résultats ont montré une déméthylation significative des loci chez les hybrides hétérotiques (Zhao *et al.* 2008). Etant donné que la méthylation / déméthylation des loci est corrélée avec l'expression des gènes, ces études suggèrent que la méthylation de l'ADN pourrait jouer un rôle dans la performance de l'hétérosis.

Une étude récente effectuée chez la plante *A. thaliana*, a montré que le croisement de deux écotypes C24 et Ler, possédant des génomes similaires, génère des hybrides qui sont 250% plus important en biomasse que les écotypes parents (figure 28 A) (Schneeberger *et al.* 2011; Groszmann *et al.* 2013). Cette même équipe a montré une implication des mécanismes épigénétiques dans le phénotype « hétérosis » obtenu. En effet, une augmentation du taux de la méthylation de l'ADN a été observée au niveau de certains loci chez les hybrides hétérozygotes. Cette augmentation de la méthylation de l'ADN est corrélée à la présence des ARN non codants dans la même région (siRNA) (figure 28 C) (Groszmann *et al.* 2011). Afin de poursuivre l'étude, une autre équipe a effectué des mutations dans la voie de biogenèse des siRNAs chez les deux écotypes parentaux. Les hybrides provenant des deux écotypes parentaux mutés ont donné naissance à des plantes ayant des biomasses similaires à celles des parents (figure 28 B) (Shen *et al.* 2012). Cette observation pourrait être interprétée par deux façons non exclusives : (i) L'épigénétique serait impliquée dans le phénotype « hétérosis » obtenu. (ii) Le croisement a impliqué deux lignées de plantes mutées pour un gène, donc dans cette situation nous pourrions avoir des effets phénotypiques liés à ces mutations.

Les hybrides dans les interactions hôte / parasite

Dans les interactions hôte / parasite, la majorité des études sont consacrées à effectuer des croisements inter et/ou intra-spécifiques des hôtes, plutôt que des parasites. Un bon nombre de ces études ont montré un phénomène d'hétérosis chez les hybrides hôtes. Phénomène qui est traduit par une meilleure résistance de l'hôte hybride F1 vis-à-vis du parasite (Moullia *et al.* 1995; Jackson & Tinsley 2003). Citons par exemple, le croisement de deux espèces différentes de souris *Mus domesticus* and

Mus musculus, a montré que la génération souris F1 est plus résistante au parasite intestinal *Aspiculuris tetraptera* que les souris parentales. En effet, ils ont trouvé beaucoup moins de vers parasites (1 à 12 vers / souris) qui se sont développés dans les souris hybrides, que dans les espèces parentales de souris (21 à 72 vers / souris) (Mouliat *et al.* 1995).

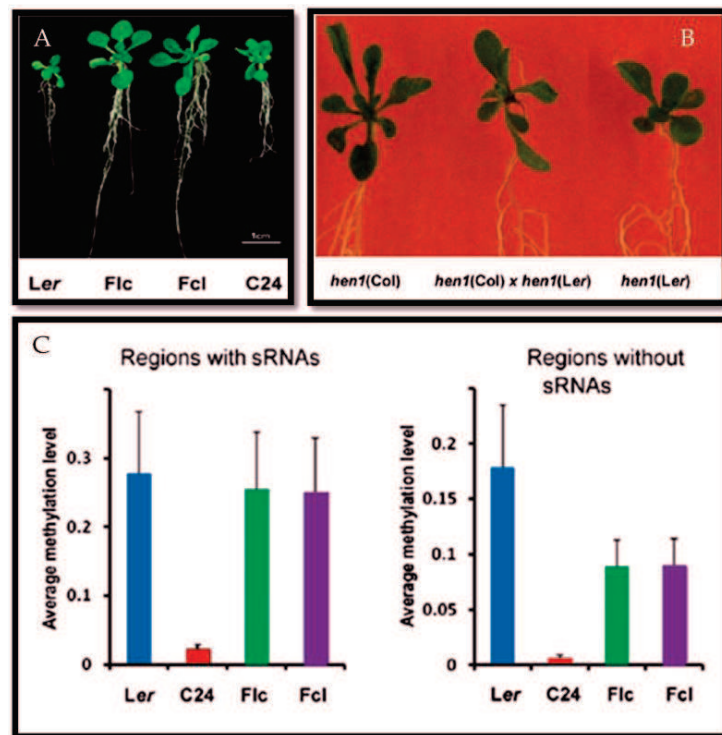


Figure 28 : Les perspectives de l'implication des mécanismes épigénétiques dans l'hétérosis chez *A. thaliana*. (A) Les phénotypes hétérotiques des plantes hybrides F1. C24 / Ler sont les lignées parentales et Fcl / Flc sont les hybrides obtenus par des croisements dans les deux sens du sexe. (B) Le compromis de l'hétérosis chez les hybrides F1 provenant de deux lignées parentales (Col / Ler) mutées. (C) Les analyses des ARN non codants dans les régions différentiellement méthylées chez les lignées parentales (Ler / C24) (Shen *et al.* 2012).

En ce qui concerne les parasites (à part les schistosomes), les études des hybridations sont rares. En effet, les vers parasites pluricellulaires rassemblent les espèces appartenant aux helminthes, qui à leur tour sont divisés en plathelminthes et némathelminthes. La majorité des espèces parasites appartenant aux helminthes sont connues pour être des hermaphrodites (Loker & Brant 2006). Ceci serait une raison principale limitante pour l'étude des hybrides.

Bien que les études des hybrides parasites soient rares, une étude récente effectuée chez le nématode *Meloidogyne incognita*, parasite dioïque des plantes, a démontré que les espèces les plus pathogènes de ce nématode proviennent à partir des croisements interspécifiques impliquant différentes espèces de *Meloidogyne* (Lunt *et al.* 2014).

Parmi les helminthes, les schistosomes trématodes sont dioïques et donc passent par une reproduction sexuée.

Les hybrides de *Schistosoma spp.*

Dans le cas des schistosomes, l'hybridation peut avoir des impacts à plus grande échelle touchant l'épidémiologie de la maladie (Huyse *et al.* 2009).

L'hybridation chez *Schistosoma spp.* est un phénomène qui a été largement observé et étudié jusqu'à l'heure actuelle. Les croisements de différentes espèces ou souches de schistosomes ont été observés en milieu naturel, mais également par des expérimentations au laboratoire. Différents exemples sont cités dans le tableau 4.

Le phénotype des hybrides de *Schistosoma spp.*

Les phénotypes observés après les croisements des schistosomes ont été essentiellement basés sur les traits de vie des hybrides. La viabilité des hybrides, la fécondité, la morphologie des œufs, l'infectivité des hôtes et la capacité d'infester différentes espèces d'hôtes intermédiaires ont été surtout pris en compte. Cependant la transmission des phénotypes à travers les générations ainsi que l'origine moléculaire de ces phénotypes n'ont pas été évoqués jusqu'à présent.

Différents cas d'hétérosis ou vigueur hybride ont été étudiés, tels que le croisement interspécifique *S. haematobium* x *S. matthei* (Wright & Ross 1980), qui a montré une augmentation de la production des œufs par femelle et une infectivité élevée pour les deux espèces (ou variétés d'hôtes intermédiaires).

Egalement, le croisement des deux différentes souches guadeloupéenne (des Caraïbes) et africaine (du Cameroun) a montré que les descendants sont plus adaptés sur leurs hôtes intermédiaires et leur hôte définitif aussi (Imbert-Establet & Combes 1986). Au-delà de ces résultats, aucune étude phénotypique moléculaire n'a été faite.

Finalement, le croisement des deux souches de *S. mansoni* (*Sm* BRE x *Sm* GH2), a montré aussi un phénomène d'hétérosis qui est traduit par la virulence de nos parasites, sur les hamsters mais aussi sur les mollusques (surtout pour la génération F2).

Espèce 1	Espèce 2	Hybridation	Fitness	Lieu	Référence
<i>S. intercalatum</i> ♀	<i>S. haematobium</i> ♂	Naturelle	Viabilité élevée	Cameroun	Wright et al 1974 Southgate 1976 Mutani 1985
<i>S. haematobium</i> ♀	<i>S. intercalatum</i> ♂	Naturelle	Viabilité réduite	Cameroun	Wright et al 1974 Southgate 1976
<i>S. intercalatum</i>	<i>S. haematobium</i>	Naturelle	viable	Egypte	Wright et al 1974 Southgate 1976
<i>S. haematobium</i>	<i>S. mattheei</i>	Naturelle	viable	Afrique du sud	Pitchford 1961
<i>S. bovis</i>	<i>S. haematobium</i>	Naturelle	Viabilité élevée	Sénégal	Huyse et al 2009
<i>S. bovis</i>	<i>S. curassoni</i>	Naturelle	Viable	Niger	Brémond 1990
<i>S. bovis</i>	<i>S. mattheei</i>	expérimentale	Viabilité réduite	ND	Taylor 1970
<i>S. mattheei</i>	<i>S. haematobium</i>	expérimentale	Viable	ND	Taylor 1970
<i>S. bovis</i>	<i>S. haematobium</i>	expérimentale	Viable	ND	Taylor 1970
<i>S. mansoni</i>	<i>S. rodhaini</i>	expérimentale	Viable	ND	LeRoux 1954 Théron 1989
<i>S. mansoni</i> ♀	<i>S. mattheei</i> ♂	expérimentale	Sterile	ND	Taylor 1970
<i>S. mattheei</i> ♀	<i>S. mansoni</i> ♂	expérimentale	Viabilité réduite	ND	Taylor 1970
<i>S. mansoni</i>	<i>S. bovis</i>	expérimentale	Viabilité réduite	ND	Taylor 1970

Tableau 4 : Exemples de croisements interspécifiques entre les différentes espèces de *Schistosoma* qui ont été étudiés après hybridation naturelle ou expérimentale. Les croisements où les symboles (♀ et ♂) sont absents, signifient que les phénotypes obtenus ne sont pas lié au sexe (ND : pour des données non déterminées) (LeRoux 1954; Pitchford 1961; Taylor 1970; Wright 1974; Southgate *et al.* 1976; Mutani *et al.* 1985; Théron 1989; Brémond *et al.* 1990; Huyse *et al.* 2009).

L'augmentation de la compatibilité avec l'hôte intermédiaire *B. glabrata*.

Au niveau de l'hôte intermédiaire, nous avons observé une augmentation de la compatibilité traduite par l'augmentation des prévalences et des intensités du parasite (Publication N°2).

L'objectif principal de la thèse c'était de déterminer le poids relatif de la génétique et de l'épigénétique. La contribution de l'épigénétique a été déterminée par la technique d'immunoprécipitation de la chromatine (ChIP) effectuée pour toutes les générations depuis la F0 jusqu'à la F3. Nous avons focalisé les analyses sur un marqueur de compatibilité, le gène *SmPoMuc 3.1 (r1r2)* avec un autre marqueur codant une enzyme métalloprotéase.

Nous avons montré également une augmentation de la compatibilité des parasites avec les mollusques en provoquant des épimutations. En effet, en traitant des miracidia incompatibles avec une souche de mollusque, à la TSA, ceci a augmenté significativement les prévalences et les intensités des parasites.

Bien que l'épigénétique semblerait jouer le rôle primordial dans la compatibilité du parasite avec le mollusque, le rôle des mécanismes génétiques n'est pas exclu.

Suite à notre objectif principal, notre but ici c'était de chercher un ou plusieurs loci dans le génome du parasite qui serait (aient) impliqué(s) dans la compatibilité avec l'hôte.

Existent-ils des gènes qui seraient directement impliqués dans la résistance du parasite ?

Existent-ils des gènes régulateurs en *cis* ou en *trans* de la structure chromatinienne d'autres gènes impliqués dans la compatibilité, notamment les *SmPoMucs* ?

Pour répondre à ces questions, nous avons initié une collaboration avec *Timothy Anderson* du *Texas Biomedical Research Institute*. Cette équipe possède tous les outils et les mis au point de la méthode *extreme Quantitative Trait Locus (X-QTL)* sur *S. mansoni*. Cette méthode consiste à déterminer un ou plusieurs gènes responsables

d'un phénotype particulier (compatibilité, comportement, résistance...) (Chevalier *et al.* 2014).

Dans leur équipe, ils ont mis au point une nouvelle stratégie pour identifier les QTL qui est basée sur trois étapes :

1. Sélection d'un pool d'individus par rapport au phénotype cherché.
2. Génotypage du pool d'individus à l'aide de la technique NGS ou pool-seq. Le séquençage massif est effectué sur l'exome de *S. mansoni* et non pas tout le génome. En effet, le génome fait en total 369 Mb (Berriman *et al.* 2009) avec 48% de séquence répétées (Lepesant *et al.* 2012), alors que l'exome fait 15 Mb environ (~20% du génome) (Chevalier *et al.* 2014). Donc, la taille de l'exome est plus réduite par rapport au génome (au total : ~68000 exons). Pour chacun des exons, ils ont désigné une sonde d'ARN biotinylée de 120 pb (Chevalier *et al.* 2014).
3. Analyse et identification des régions du génome qui sont impliquées dans ce phénotype par comparaison des fréquences alléliques.

Revenant au schéma expérimental de la figure 20 - Chapitre II :

Les échantillons d'ADNg des populations de cercaires générées au cours des croisements des deux souches du parasite, ont été envoyés afin d'être analysés par X-QTL. Ces populations clonales de cercaires sont indiquées par le mot : Génotypage et flèche grise (figure 20).

Pour l'analyse, 30 échantillons (pool) ont été envoyés :

- 2 parentaux *Sm* BRE et *Sm* GH2
- 1 *Sm* F1
- 27 *Sm* F2 : 17 populations de cercaires issues chacune d'un mollusque *Bg* GUA et 10 populations issues chacune d'un mollusque *Bg* BRE.

Les analyses étaient consacrées surtout sur les populations de cercaires de la F2. En effet, 17 populations sont passées sur le mollusque *Bg* GUA et 10 populations sont passées sur le mollusque *Bg* BRE.

Les résultats de X-QTL obtenus (en aout 2014) ont montré que deux régions du génome de *S. mansoni* seraient potentiellement intéressantes : L'une région sur le chromosome 6 et l'autre sur le chromosome Z. Ce résultat a permis de suggérer que ces deux régions pourraient être impliquées dans la compatibilité avec les mollusques *Bg* BRE et *Bg* GUA. Cependant, la correction de Bonferroni a montré que ces deux pics correspondants observés, ne sont pas significativement différents par rapport aux autres pics obtenus. Ces résultats restent préliminaires à l'heure actuelle. Pour ces analyses, il manque surtout les réplicats biologiques (il faudra augmenter le nombre de populations de cercaires pour chacune des conditions) afin de pouvoir en tirer des conclusions.

Nous n'avons pas obtenu de pic significatif, ceci pourrait être interprété de deux manières différentes :

Il n'existerait pas de locus particulier impliqué dans la compatibilité. Ceci serait en faveur de l'hypothèse que des mécanismes épigénétiques seraient une source d'une adaptation rapide. Ils permettent la mise en place du « epigenetic mutation system first » (Klironomos *et al.* 2013).

Ou bien, les régions du chromosome 6 et/ou du chromosome Z contiendraient des loci intéressants et que le nombre insuffisant de populations a rendu les pics correspondants non significatifs. Ces deux régions restent à investiguer.

IV.2. Les modifications épigénétiques chez *B. glabrata*

Les résultats obtenus par la MS-AFLP ont montré que les variations génétiques sont très importantes par rapport aux variations épigénétiques entre les différents individus *B. glabrata* d'une même population. Ce qui rend cette technique non adaptée pour l'étude des variations épigénétiques au niveau populationnel chez l'espèce *B. glabrata*. Par contre, cette technique a montré son efficacité au niveau des études de variation génétique chez le mollusque. En effet, nous avons pu démontrer que les deux différentes souches de *B. glabrata*, *Bg* BRE et *Bg* GUA, sont génétiquement différentes. Ceci est corrélé par le fait que ces deux souches proviennent de deux origines géographiques différentes.

De plus, ces analyses ont montré que la diversité génétique entre les individus d'une même population est faible. A l'exception des individus appartenant à la population des *Bg* GUAEni exposés non infestés, qui sont les plus diversifiés. Ces résultats pourraient être intéressants pour répondre aux hypothèses proposées sur les diversifications génétiques inter-individuelles ou même intra-individuelles. Des études ont montré que le taux d'infection du mollusque *B. alexandrina* par *S. mansoni* est dépendant de la variabilité génétique des mollusques (Mohamed *et al.* 2012) : cas de diversification génétique inter-individuelle. D'autres études ont montré que les FREPs, immunoglobulines diversifiées du mollusque qui sont essentielles à l'interaction avec le parasite, sont hautement diversifiées au sein d'un même individu (Zhang *et al.* 2004) : cas de diversification génétique intra-individuelle.

Les résultats du chapitre III (Publication N°3) ont montré que le génome de *B. glabrata* est faiblement méthylé, 2% de cytosines sont méthylés, ceci pourrait expliquer aussi la difficulté de détecter des variations épigénétiques par MS-AFLP.

Afin de répondre à la question de cette partie : si l'épigénome du mollusque varie en fonction de l'infestation par le parasite, il serait essentiel de passer à une autre espèce de mollusque qui serait génétiquement moins diversifiée.

Les deux plus importants hôtes intermédiaires pour *S. mansoni* sont *B. glabrata*, l'espèce du nouveau monde, et *B. pfeifferi*, l'espèce de l'ancien monde (origine africaine). Cependant, les séquences de l'ADN mitochondrial de *B. pfeifferi* sont (4x) moins diversifiées que celles de *B. glabrata* (DeJong *et al.* 2003).

Comme *B. pfeifferi* est élevée dans notre laboratoire aussi, elle a été l'espèce de choix pour poursuivre l'étude des variations épigénétiques. Effectivement, nous avons montré par MS-AFLP qu'il n'existe quasiment pas de variations génétiques entre les différents individus de *B. pfeifferi* testés (données non publiées).

Pour initier l'étude de méthylation de l'ADN sur *B. pfeifferi*, nous avons commencé par la détermination du taux de 5mC et 5hmC à l'échelle globale du génome. La même démarche décrite pour *B. glabrata* dans la publication N°3, a été suivie. Les

résultats ont montré que 0.45% des cytosines totales sont méthylés dans le génome de *B. pfeifferi* (tableau 5).

Espèces / souche	% en 5mC	% en 5hmC
<i>B. glabrata</i> BRE	2%	0.0009 %
<i>B. glabrata</i> GUA	2%	0.0009 %
<i>B. pfeifferi</i>	0.45 %	0.00

Tableau 5 : Les taux de 5mC et 5hmC chez *B. pfeifferi*. Les résultats sont obtenus par LC/MS.

Ces résultats montrent que le taux de méthylation varie fortement entre les deux espèces de mollusques. Ceci est de l'ordre de (4x) moins de cytosines totales méthylées chez *B. pfeifferi* que chez *B. glabrata*.

Cette différence de taux global de méthylation des cytosines entre les deux espèces de mollusques est curieuse. Les analyses phylogénétiques ont montré que *B. glabrata* est la seule espèce sœur de toutes les espèces africaines de *Biomphalaria* (Woodruff & Mulvey 1997; DeJong *et al.* 2003). En se basant sur des analyses de séquences de l'ADN mitochondrial et nucléaire des différentes espèces, ces études ont suggéré que l'ancêtre de *B. glabrata* a tout d'abord colonisé l'Afrique. Il a donné ensuite naissance à toutes les autres espèces africaines incluant *B. pfeifferi* (Woodruff & Mulvey 1997; Campbell *et al.* 2000). Etant donné que *B. glabrata* est l'espèce du nouveau monde, ce mollusque a dû s'adapter à des nouvelles conditions géographiques, nouveau climat et nouvel écosystème.

En parallèle, les études de phylogénie montrent que le parasite *S. mansoni* a été transférée récemment en Amérique du sud (Zhou *et al.* 2008). *S. mansoni* a montré une forte capacité d'adaptation sur différentes espèces de *Biomphalaria* du nouveau monde (Morgan *et al.* 2001). A son tour, *B. glabrata* s'est adaptée à l'espèce de parasite *S. mansoni*.

L'adaptation du mollusque à toutes les conditions de vie dans le nouveau monde a nécessité des variations phénotypiques. D'un point de vue évolutif, nous suggérons que des mécanismes épigénétiques chez *B. glabrata*, notamment la méthylation de l'ADN, pourraient être « une » des sources de son adaptation au nouveau monde.

Ceci reste hypothétique à l'heure actuelle, pour cela il serait essentiel de mener des études plus approfondies qui permettraient une comparaison globale de tous les sites méthylés dans les deux génomes de *B. glabrata* et *B. pfeifferi*. Nous proposons une approche de Bisulfite sequencing (BS-seq) afin de comparer les méthylomes des deux espèces.

Sachant que chez les invertébrés, la méthylation de l'ADN a lieu préférentiellement dans les unités transcriptionnelles, les approches de séquençage massif (par RNA-seq) apporteraient des informations supplémentaires sur la méthylation / expression des gènes. Les résultats permettraient d'effectuer des liens entre les variations épigénétiques, l'expression des gènes et leurs fonctions et implications dans les processus adaptatif.

Chapitre V

Conclusions et perspectives

Chapitre V : Conclusions et perspectives

Comprendre les origines de la variabilité phénotypique est très important pour la biologie évolutive, incluant les sciences médicales, l'agriculture et autres branchements de la biologie. Les études évolutives récentes montrent de plus en plus l'importance de la variabilité épigénétique dans l'adaptation rapide surtout dans les environnements fluctuants. L'évolution adaptative est un processus basé sur la variabilité phénotypique et son héritabilité au cours des générations (Darwin 1868; Cuénot 1925). Les biologistes étaient habitués à la réflexion que la diversité génétique est la seule source des variations phénotypiques. Cependant, la variabilité phénotypique pourrait être observée dans le cas où des individus sont génétiquement similaires (Norouzitallab *et al.* 2014). Au cours des dernières décennies, les modifications épigénétiques sont considérées comme des mécanismes alternatifs qui induisent des changements phénotypiques sans modification du génotype. Des études théoriques considèrent que les modifications épigénétiques pourraient constituer la première étape dans l'évolution adaptative par l'hypothèse « epigenetic mutation system first » (Pal & Miklos 1999; Klironomos *et al.* 2013).

Afin d'évaluer les origines de l'adaptation rapide, les systèmes hôte / parasite constituent de bons modèles pour cette étude. En effet, les effets du parasite sur l'hôte sont inséparables des effets de l'hôte sur le parasite. Les deux partenaires forment un système d'interaction très dynamique (Van Valen 1974). La coadaptation de l'un sur l'autre s'effectuerait souvent dans un intervalle de temps qui est court (en fonction des espèces). Dans ce sens, les modifications épigénétiques pourraient assurer une source accessoire pour la mise en place d'une variabilité phénotypique, relativement rapide, réversible et soumise aux pressions de sélection des deux partenaires (Gómez-Díaz *et al.* 2012).

Les deux volets principaux de cette thèse : (i) L'étude de la contribution des mécanismes épigénétiques dans l'expression des variants phénotypiques adaptatifs chez le parasite *S. mansoni* en se focalisant sur les *SmPoMucs*, marqueurs

moléculaires qui semblent être essentiels pour la compatibilité. Et (ii) La mise en évidence de la méthylation de l'ADN chez son hôte intermédiaire *B. glabrata*.

Chez le parasite *S. mansoni*, nous avons démontré que les modifications chimiques de la chromatine autour des promoteurs jouent un rôle majeur dans la régulation de la transcription de certains groupes de *SmPoMucs*.

Les croisements des deux différentes souches de parasite *Sm* BRE et *Sm* GH2, a donné naissance à des hybrides à compatibilité élevée. Cette compatibilité accrue est qualifiée par le phénomène d'hétérosis ou vigueur hybride. Elle a été observée chez l'hôte intermédiaire *B. glabrata* par des taux élevés de prévalences et des intensités, et chez l'hôte définitif (hamsters) par l'augmentation de nombre d'œufs par foie et te taux de mortalité. Au niveau moléculaire, nous avons observé l'expression de nouveaux variants phénotypiques des marqueurs de compatibilité, *SmPoMucs*. Ce phénomène d'hétérosis a été souvent expliqué par des mécanismes de diversification qui sont basés surtout sur des facteurs génétiques. Cependant au cours de ce travail, nous avons montré que les modifications de la structure chromatinienne pourraient jouer un rôle important aussi, notamment au niveau du marqueur, *SmPoMuc* 3.1 (r1r2). En effet, l'expression de ce marqueur moléculaire est régulée surtout par des modifications épigénétiques.

Afin de valider l'implication des modifications épigénétiques sur l'expression de *SmPoMuc* 3.1 (r1r2), nous avons voulu modifier la structure chromatinienne de son promoteur. L'approche idéale serait de modifier directement la chromatine au niveau de ce gène d'une manière ciblée, mais ceci est infaisable à l'heure actuelle. Par contre, nous avons effectué des modifications épigénétiques à l'échelle globale de l'épigénome à l'aide d'un agent épimutagène, la TSA. En traitant des miracidia de différentes souches de parasites à la TSA, nous avons observé également une augmentation de la compatibilité (prévalences et intensités) du parasite vis-à-vis du mollusque et l'expression de nouveaux variants *SmPoMucs*. Ces résultats ont confirmé l'implication des modifications épigénétiques dans le contrôle de l'expression des *SmPoMucs* ainsi que dans l'apparition de nouveaux variants phénotypiques. A partir du moment où ces phénotypes seraient essentiels pour le

parasite (telle que sa compatibilité avec son hôte), nous pourrions considérer que ces phénotypes seraient adaptatifs.

Ces résultats sont en faveur des théories proposées sur l'intervention des mécanismes épigénétiques en premier lieu, dans la mise en place des variants phénotypiques adaptatifs (Klironomos *et al.* 2013).

Afin de poursuivre le travail du premier volet de la thèse :

D'après la théorie du « *Dual inheritance system* », les mécanismes génétiques et les mécanismes épigénétiques contribuent ensemble à l'expression du phénotype sous influence de l'environnement (Maynard Smith 1990; Danchin *et al.* 2011). Mon travail de thèse était focalisé sur la contribution de l'épigénétique dans l'expression des variants phénotypiques de *SmPoMucs*.

Nos résultats ont montré par des approches ciblées que l'expression du gène *SmPoMuc 3.1 (r1r2)* est régulée par des modifications épigénétiques, cependant ce résultat n'exclut pas un rôle d'autres facteurs génétiques et / ou épigénétiques au niveau des autres régions du génome du parasite.

Afin de chercher d'éventuels marqueurs génétiques impliqués dans la compatibilité du parasite vis-à-vis du mollusque, nous avons collaboré avec *Timothy Anderson*. Les résultats obtenus par *extreme Quantitative Trait Locus (X-QTL)* (expliquée dans le chapitre IV) sur les générations hybrides du parasite, n'ont pas déterminé la présence d'un marqueur génétique significativement intéressant. Par contre, ils ont ouvert des pistes importantes sur la recherche d'éventuels facteurs impliqués dans les interactions parasite / mollusque, la résistance ou bien des facteurs liés aux effets parentaux (de souches).

C'était une étude pilote qui a montré essentiellement la faisabilité de l'approche. Ces résultats pourraient supporter deux interprétations différentes : (i) Effectivement, il n'existe pas de locus responsable du phénotype de la compatibilité du parasite vis-à-vis du mollusque et ce serait la structure chromatinienne qui est importante au niveau du génome entier. (ii) Le nombre de réplicats biologiques des populations de

parasites étudiées était insuffisant pour montrer si le résultat obtenu est significatif ou pas. Pour cela, une nouvelle collaboration est actuellement lancée par différents membres du laboratoire 2EI, afin de poursuivre cette étude mais à une échelle plus large.

Jusqu'à présent, nous suggérons que les *SmPoMucs* joueraient un rôle clé dans la compatibilité, cependant leur rôle exact est encore énigmatique. L'application de la technique *Knock down* des *SmPoMucs* par RNAi a été mise au point. À l'heure actuelle, cette approche est poursuivie jusqu'à l'étude phénotypique, dans le but d'observer la capacité des miracidia traités à infester. Cette application reste difficile dans le sens où les *SmPoMucs* appartiennent à une famille multigénique. Ils sont divisés en trois groupes fonctionnels, ceci pourrait leur attribuer des fonctions différentes.

Pour rester dans les approches moléculaires, nous avons utilisé au cours de ce travail un autre marqueur, la métalloprotéase *Smp_171100*, comme un contrôle pour le marqueur *SmPoMuc 3.1 (r1r2)*. Ce marqueur, comme *SmPoMuc 3.1 (r1r2)*, est différentiellement exprimé entre les deux souches du parasite *Sm BRE* et *Sm GH2* ainsi que sa structure chromatinienne est différente entre les deux souches. J'ai effectué récemment une étude de transcription qui a montré une expression spécifique de ce gène au stade de miracidia. Etant donné que les métalloprotéases joueraient un rôle dans l'immunosuppression de leurs hôtes (Bland *et al.* 2008), l'approche fonctionnelle de ce gène sera envisageable. En effet, ce gène pourrait être un élément dans le cœur de l'interaction aussi, ou bien il pourrait ouvrir des pistes sur d'autres éléments de l'interaction également. L'avantage de ce gène par rapport aux *SmPoMucs* c'est qu'il est présent en copie unique dans le génome et par conséquent, les approches moléculaires fonctionnelles seraient plus faciles à être appliquées.

Le traitement des parasites à la TSA a montré une apparition de nouveaux phénotypes, qui sont traduits par l'augmentation de la compatibilité vis-à-vis du mollusque et l'expression de nouveaux variants moléculaires. Afin de tester

l'héritabilité de ces variants au cours des générations, cette approche sera effectuée dans le cadre de la thèse de Benoit Aliaga.

Finalement, en ce qui concerne l'héritabilité des modifications épigénétiques, nous avons montré au cours de ce travail deux différents types de transmission des modifications épigénétiques. (i) La transmission de la structure chromatinienne a suivi les lois mendéliennes pour le gène codant la métalloprotéase (où l'épiallèle a co-ségré avec l'allèle). Cependant, (ii) La transmission n'a pas suivi les lois mendéliennes pour le gène codant la *SmPoMuc 3.1 (r1r2)* (où l'épiallèle a été transmis d'une manière indépendante de l'allèle). Ce qui était remarquable, c'est que pour les deux types de transmission : l'épigénotype de la souche brésilienne (*Sm BRE*) qui a été transmis des parents jusqu'à la génération F3. Cette observation a ouvert une question importante concernant le type de transmission des modifications épigénétiques.

Est-ce que c'est l'épigénome entier de la souche *Sm BRE* qui est transmis en bloc des parents aux descendants (par effet parental)? Ou bien l'épigénome de *S. mansoni* se transmet d'une façon aléatoire (mendélienne et / ou non mendélienne) en fonction des loci ?

Afin de répondre à cette question, il serait important d'envisager du ChIP suivi par du séquençage massif ou ChIP-seq. Ceci permettra l'étude comparative globale des modifications épigénétiques à l'échelle des épigénomes de la F0 à la F3.

Chez l'hôte *B. glabrata*, nous avons montré que les gènes codant la machinerie enzymatique de la méthylation de l'ADN sont présents dans le génome. A l'échelle globale du génome du mollusque, 2% des cytosines sont méthylées. Chez *B. glabrata*, comme chez la plupart des invertébrés étudiés jusqu'à présent, la méthylation de l'ADN est de type mosaïque. En effet, les transcripts sont divisés en deux lots de gènes : des gènes qui sont fortement méthylés et d'autres gènes qui sont faiblement méthylés. Les gènes fortement méthylés appartiennent éventuellement à des gènes de ménage. Finalement, nous étions les premiers à décrire la présence de la

méthylation de l'ADN chez *B. glabrata*. Ceci ouvre une voie de recherche très large autour de cette modification épigénétique et son rôle effectué chez le mollusque.

Afin de poursuivre le travail du deuxième volet de la thèse :

Existe-t-il des variations de méthylation de l'ADN entre des mollusques infestés par le parasite et d'autres mollusques non infestés ?

Afin de répondre à cette question, des approches globales de séquençage des méthylomes par Bisulfite-sequencing (BS-seq) sont en cours par mon équipe.

La méthylation de l'ADN serait-ils une des sources des variants phénotypiques exprimés chez le mollusque comme déjà démontré chez le parasite ?

Il serait important de réaliser des approches de traitement des mollusques à des agents épimutagènes. Afin de provoquer des épimutations, un inhibiteur des ADN méthyltransférases (5-azacytidine) ou même l'inhibiteur des histones désacétylases, TSA, pourrait être utilisés. Ces variations épigénétiques visent à étudier les variations phénotypiques du mollusque vis-à-vis du parasite cette fois. Les phénotypes de compatibilité et d'expression des variants moléculaires (FREPs par exemple), seront envisagés.

Chapitre VI

Références bibliographiques

Chapitre VI : Références bibliographiques

- Adema, C., Loker, E., Fried, B. & Graczyk, T. (1997a). Specificity and immunobiology of larval digenean-snail associations. *Advances in trematode biology*, 229-263.
- Adema, C.M., Hertel, L.A., Miller, R.D. & Loker, E.S. (1997b). A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 8691-8696.
- Agrawal, A.A. (2001). Ecology - Phenotypic plasticity in the interactions and evolution of species. *Science*, 294, 321-326.
- Agrawal, A.A., Laforsch, C. & Tollrian, R. (1999). Transgenerational induction of defences in animals and plants. *Nature*, 401, 60-63.
- Anderson, R. & May, R. (1982). Coevolution of hosts and parasites. *Parasitology*, 85, 411-426.
- Antequera, F. & Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences*, 90, 11995-11999.
- Arican-Goktas, H.D., Ittiprasert, W., Bridger, J.M. & Knight, M. (2014). Differential spatial repositioning of activated genes in *Biomphalaria glabrata* snails infected with *Schistosoma mansoni*. *PLoS neglected tropical diseases*, 8, e3013.
- Azzi, A., Cosseau, C. & Grunau, C. (2009). *Schistosoma mansoni*: Developmental arrest of miracidia treated with histone deacetylase inhibitors. *Experimental Parasitology*, 121, 288-291.
- Bachellerie, J.-P., Cavaillé, J. & Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, 84, 775-790.
- Bagheri, N. & Jelodar, N.B. (2010). Heterosis and combining ability analysis for yield and related-yield traits in hybrid rice. *International Journal of Biology*, 2, P222.
- Baranwal, V.K., Mikkilineni, V., Zehr, U.B., Tyagi, A.K. & Kapoor, S. (2012). Heterosis: emerging ideas about hybrid vigour. In: *Journal of Experimental Botany*. Soc Experiment Biol, pp. 6309-6314.
- Barth, T.K. & Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. *Trends in biochemical sciences*, 35, 618-626.
- Basch, P.F. (1975). An interpretation of snail-trematode infection rates: specificity based on concordance of compatible phenotypes. *International journal for parasitology*, 5, 449-452.
- Beltran, S., Galinier, R., Allienne, J. & Boissier, J. (2008). Cheap, rapid and efficient DNA extraction method to perform multilocus microsatellite genotyping on all *Schistosoma mansoni* stages. *Memorias do Instituto Oswaldo Cruz*, 103, 501-503.
- Berger, S.L. (2007). The complex language of chromatin regulation during transcription. *Nature*, 447, 407-412.

- Bernstein, E. & Allis, C.D. (2005). RNA meets chromatin. *Genes & development*, 19, 1635-1655.
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C. *et al.* (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, 309, 416-422.
- Berriman, M., Haas, B.J., LoVerde, P.T., Wilson, R.A., Dillon, G.P., Cerqueira, G.C. *et al.* (2009). The genome of the blood fluke *Schistosoma mansoni*. In: *Nature* Nature Publishing Group, pp. 352-358.
- Berry, A., Moné, H., Iriart, X., Mouahid, G., Abbo, O., Boissier, J. *et al.* (2014). Schistosomiasis Haematobium, Corsica, France *Emerg Infect Dis* 20.
- Bird, A. (2007). Perceptions of epigenetics. In: *Nature* Nature Publishing Group, pp. 396-398.
- Bird, A.P. (1985). CpG-rich islands and the function of DNA methylation. *Nature*, 321, 209-213.
- Bland, N., Pinney, J., Thomas, J., Turner, A. & Isaac, R.E. (2008). Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. In: *BMC evolutionary biology*. BioMed Central Ltd, p. 16.
- Bonin, A., Ehrich, D. & Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular ecology*, 16, 3737-3758.
- Bossdorf, O., Richards, C.L. & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters*, 11, 106-115.
- Brémond, P., Mouchet, F., Chevallier, P., Sellin, E., Véra, C. & Sellin, B. (1990). Flux génique entre *Schistosoma bovis* et *S. curassoni* au Niger. In: *Bulletin de la Société Française de Parasitologie*.
- Bull, J. (1994). Perspective: virulence. *Evolution*, 1423-1437.
- Burdge, G.C. & Lillycrop, K.A. (2010). Nutrition, epigenetics, and developmental plasticity: implications for understanding human disease. *Annual review of nutrition*, 30, 315-339.
- Cai, P., Piao, X., Hao, L., Liu, S., Hou, N., Wang, H. *et al.* (2013). A Deep Analysis of the Small Non-Coding RNA Population in *Schistosoma japonicum* Eggs. In: *PloS one*. Public Library of Science, p. e64003.
- Campbell, G., Jones, C.S., Lockyer, A.E., Hughes, S., Brown, D., Noble, L.R. *et al.* (2000). Molecular evidence supports an African affinity of the Neotropical freshwater gastropod, *Biomphalaria glabrata*, Say 1818, an intermediate host for *Schistosoma mansoni*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267, 2351-2358.
- Champagne, F.A. (2008). Epigenetic mechanisms and the transgenerational effects of maternal care. *Frontiers in neuroendocrinology*, 29, 386-397.

- Cheng, X. & Blumenthal, R.M. (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure*, 16, 341-350.
- Chevalier, F.D., Valentim, C.L., LoVerde, P.T. & Anderson, T.J. (2014). Efficient linkage mapping using exome capture and extreme QTL in schistosome parasites. *BMC genomics*, 15, 617.
- Chitsulo, L., Engels, D., Montresor, A. & Savioli, L. (2000). The global status of schistosomiasis and its control. *Acta tropica*, 77, 41-51.
- Combes, C. (1988). L'ADAPTATION.
- Comstock, R. & Moll, R.H. (1963). Genotype-environment interactions. *Statistical genetics and plant breeding*, 164-196.
- Cosseau, C., Azzi, A., Rognon, A., Boissier, J., Gourbiere, S., Roger, E. *et al.* (2010). Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* - a possible kick-off for adaptive host/parasite evolution. *Oikos*, 119, 669-678.
- Cosseau, C., Azzi, A., Smith, K., Freitag, M., Mitta, G. & Grunau, C. (2009). Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Molecular and Biochemical Parasitology*, 166, 70-76.
- Costa, F.F. (2005). Non-coding RNAs: new players in eukaryotic biology. *Gene*, 357, 83-94.
- Costa, F.F. (2008). Non-coding RNAs, epigenetics and complexity. *Gene*, 410, 9-17.
- Coulondre, C., Miller, J.H., Farabaugh, P.J. & Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274, 775-780.
- Cuénot, L. (1925). *L'adaptation*. G. Doin.
- Dam, T.K. & Brewer, C.F. (2010). Lectins as pattern recognition molecules: The effects of epitope density in innate immunity*. *Glycobiology*, 20, 270-279.
- Danchin, É., Charmantier, A., Champagne, F.A., Mesoudi, A., Pujol, B. & Blanchet, S. (2011). Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nature Reviews Genetics*, 12, 475-486.
- Danchin, É. & Wagner, R.H. (2010). Inclusive heritability: combining genetic and non-genetic information to study animal behavior and culture. *Oikos*, 119, 210-218.
- Darwin, C. (1859). *L'origine des espèces*.
- Darwin, C. (1868). *The variation of animals and plants under domestication*. John Murray.
- De Vries, H. (1904). *Species and varieties, their origin by mutation*. Open court publishing Company.
- DeJong, R., Morgan, J., Wilson, W., Al-Jaser, M., Appleton, C., Coulibaly, G. *et al.* (2003). Phylogeography of *Biomphalaria glabrata* and *B. pfeifferi*, important intermediate hosts of *Schistosoma mansoni* in the New and Old World tropics. *Molecular ecology*, 12, 3041-3056.

- DeJong, R.J., Morgan, J.A., Paraense, W.L., Pointier, J.-P., Amarista, M., Ayeh-Kumi, P.F. *et al.* (2001). Evolutionary relationships and biogeography of Biomphalaria (Gastropoda: Planorbidae) with implications regarding its role as host of the human bloodfluke, *Schistosoma mansoni*. *Molecular Biology and Evolution*, 18, 2225-2239.
- Devaskar, S.U. & Raychaudhuri, S. (2007). Epigenetics—a science of heritable biological adaptation. *Pediatric research*, 61, 1R-4R.
- Dias, B.G. & Ressler, K.J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature neuroscience*, 17, 89-96.
- Doby, J., Rault, B., Deblock, S. & Chabaud, A. (1965). [Snails and bilharziasis in Corsica. Distribution, frequency and biology of " *Bulinus truncatus*"]. *Annales de parasitologie humaine et comparee*, 41, 337-349.
- Doenhoff, M.J., Cioli, D. & Utzinger, J. (2008). Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Current opinion in infectious diseases*, 21, 659-667.
- Dubois, F., Caby, S., Oger, F., Cosseau, C., Capron, M., Grunau, C. *et al.* (2009). Histone deacetylase inhibitors induce apoptosis, histone hyperacetylation and up-regulation of gene transcription in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 168, 7-15.
- Duvaux-Miret, O., Stefano, G.B., Smith, E.M., Dissous, C. & Capron, A. (1992). Immunosuppression in the definitive and intermediate hosts of the human parasite *Schistosoma mansoni* by release of immunoactive neuropeptides. *Proceedings of the National Academy of Sciences*, 89, 778-781.
- Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W.M. *et al.* (2001). Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. In: *PNAS* National Acad Sciences, pp. 6209-6214.
- Egloff, R. & Fehr, J. (2011). Vérité, Widerstand, Development: At Work with / Arbeiten mit / Travailler avec Ludwik Fleck.
- Elango, N. (2008). DNA methylation and structural and functional bimodality of vertebrate promoters. In: *Molecular Biology and Evolution*. SMOE, pp. 1602-1608.
- Emmanuel, R. (2008). Bases génétiques du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata*. In: *Académie de Montpellier Université de Perpignan*
- Fantappiè, M.R., Rodrigues Pereira Gimba, E. & Rumjanek, F.D. (2001). Lack of DNA Methylation in *Schistosoma mansoni*. *Experimental parasitology*, 98, 162-166.
- Feng, S.H., Cokus, S.J., Zhang, X.Y., Chen, P.Y., Bostick, M., Goll, M.G. *et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 8689-8694.

- Fenwick, A., Savioli, L., Engels, D., Robert Bergquist, N. & Todd, M.H. (2003). Drugs for the control of parasitic diseases: current status and development in schistosomiasis. *Trends in parasitology*, 19, 509-515.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A. *et al.* (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 473, 398-402.
- Flemming (1882). *Zellsubstanz, Kern und Zelltheilung*, Leipzig.
- Flor, H.H. (1971). Current status of the gene-for-gene concept. *Annual review of phytopathology*, 9, 275-296.
- Frank, S.A. (2000). Polymorphism of attack and defense. *Trends in ecology & evolution*, 15, 167-171.
- Frésard, L., Morisson, M., Brun, J.-M., Collin, A., Pain, B., Minvielle, F. *et al.* (2013). Epigenetics and phenotypic variability: some interesting insights from birds. *Genet. Sel. Evol*, 45, 16.
- Fryxell, K.J. & Zuckerkandl, E. (2000). Cytosine deamination plays a primary role in the evolution of mammalian isochores. In: *Molecular biology and evolution*. SMOBE, pp. 1371-1383.
- Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Prados, J. *et al.* (2014). Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nature neuroscience*.
- Gavery, M. & Roberts, S. (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). In: *BMC Genomics*. BioMed Central Ltd, p. 483.
- Gavery, M.R. & Roberts, S.B. (2013). Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. *PeerJ*, 1, e215.
- Geyer, K.K., Lopez, C.M.R., Chalmers, I.W., Munshi, S.E., Truscott, M., Heald, J. *et al.* (2011). Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nature Communications*, 2.
- Glastad, K.M., Hunt, B.G., Yi, S.V. & Goodisman, M.A.D. (2011). DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol. Biol.*, 20, 553-565.
- Godfrey, K.M., Lillycrop, K.A., Burdge, G.C., Gluckman, P.D. & Hanson, M.A. (2007). Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatric research*, 61, 5R-10R.
- Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X. *et al.* (2006). Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science*, 311, 395-398.
- Gómez-Díaz, E., Jordà, M., Peinado, M. & Rivero, A. (2012). Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS pathogens*, 8, e1003007.

- Greaves, I.K., Groszmann, M., Ying, H., Taylor, J.M., Peacock, W.J. & Dennis, E.S. (2012). Trans chromosomal methylation in Arabidopsis hybrids. *Proceedings of the National Academy of Sciences*, 109, 3570-3575.
- Groszmann, M., Greaves, I.K., Albertyn, Z.I., Scofield, G.N., Peacock, W.J. & Dennis, E.S. (2011). Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proceedings of the National Academy of Sciences*, 108, 2617-2622.
- Groszmann, M., Greaves, I.K., Fujimoto, R., James Peacock, W. & Dennis, E.S. (2013). The role of epigenetics in hybrid vigour. *Trends in Genetics*, 29, 684-690.
- Gryseels, B., Mbaye, A., De Vlas, S., Stelma, F., Guisse, F., Van Lieshout, L. *et al.* (2001). Are poor responses to praziquantel for the treatment of Schistosoma mansoni infections in Senegal due to resistance? An overview of the evidence. *Tropical Medicine & International Health*, 6, 864-873.
- Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. (2006). Human schistosomiasis. *The Lancet*, 368, 1106-1118.
- Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S. & Fire, A. (2012). Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nature genetics*, 44, 157-164.
- Gutierrez, A. & Sommer, R.J. (2004). Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes Pristionchus pacificus, Caenorhabditis elegans and Caenorhabditis briggsae. *Nucleic acids research*, 32, 6388-6396.
- Hallgrímsson, B. & Hall, B.K. (2011). *Epigenetics: linking genotype and phenotype in development and evolution*. Univ of California Press.
- Han, Z., Mtango, N.R., Patel, B.G., Sapienza, C. & Latham, K.E. (2008). Hybrid vigor and transgenerational epigenetic effects on early mouse embryo phenotype. In: *Biology of REPRODUCTION Soc Study Reprod*, pp. 638-648.
- Hanington, P.C., Forys, M.A. & Loker, E.S. (2012). A somatically diversified defense factor, FREP3, is a determinant of snail resistance to schistosome infection. *PLoS neglected tropical diseases*, 6, e1591.
- Hanington, P.C. & Zhang, S.-M. (2010). The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. *Journal of innate immunity*, 3, 17-27.
- Heithoff, D.M., Sinsheimer, R.L., Low, D.A. & Mahan, M.J. (1999). An essential role for DNA adenine methylation in bacterial virulence. *Science*, 284, 967-970.
- Herrera, C.M. & Bazaga, P. (2010). Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. In: *New Phytologist*. Wiley Online Library, pp. 867-876.
- Herrera, C.M. & Bazaga, P. (2011). Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. In: *Molecular Ecology*. Wiley Online Library, pp. 1675-1688.

- Hochholdinger, F. & Hoecker, N. (2007). Towards the molecular basis of heterosis. In: *Trends in plant science*. Elsevier, pp. 427-432.
- Horsthemke, B. (2006). Epimutations in human disease. In: *DNA Methylation: Development, Genetic Disease and Cancer*. Springer, pp. 45-59.
- Huysse, T., Webster, B.L., Geldof, S., Stothard, J.R., Diaw, O.T., Polman, K. *et al.* (2009). Bidirectional introgressive hybridization between a cattle and human schistosome species. In: *Plos Pathogens* Public Library of Science, p. e1000571.
- Imbert-Establet, D. & Combes, C. (1986). *Schistosoma mansoni*: Comparison of a Caribbean and African strain and experimental crossing based on compatibility with intermediate hosts and *Rattus rattus*. In: *Experimental Parasitology*. Elsevier, pp. 210-218.
- Ismail, M., Botros, S., Metwally, A., William, S., Farghally, A., Tao, L.-F. *et al.* (1999). Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *The American journal of tropical medicine and hygiene*, 60, 932-935.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. In: *nature*. Nature Publishing Group, pp. 1129-1133.
- Jablonka, E., Lachmann, M. & Lamb, M.J. (1992). Evidence, mechanisms and models for the inheritance of acquired characters. *Journal of Theoretical Biology*, 158, 245-268.
- Jablonka, E. & Lamb, M.J. (1995). Epigenetic inheritance and evolution. *The Lamarckian Dimension*.
- Jablonka, E. & Lamb, M.J. (1999). Epigenetic Inheritance and Evolution: The Lamarckian Dimension.
- Jablonka, E., Oborny, B., Molnar, I., Kisdi, E., Hofbauer, J. & Czaran, T. (1995). The adaptive advantage of phenotypic memory in changing environments. In: *The Royal Society B*. The Royal Society, pp. 133-141.
- Jablonka, E. & Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly review of biology*, 84, 131-176.
- Jackson, J. & Tinsley, R. (2003). Parasite infectivity to hybridising host species: a link between hybrid resistance and allopolyploid speciation? *International journal for parasitology*, 33, 137-144.
- Johnston, M.O. (1992). Effects of cross and self-fertilization on progeny fitness in *Lobelia cardinalis* and *L. siphilitica*. In: *Evolution*. JSTOR, pp. 688-702.
- Julie, L. (2012). Étude de la variabilité phénotypique, sous influence environnementale, chez *Schistosoma mansoni*, parasite de l'homme :

implication de l'épigénome par une approche globale. Université de Perpignan

- Klironomos, F.D., Berg, J. & Collins, S. (2013). How epigenetic mutations can affect genetic evolution: model and mechanism. *Bioessays*, 35, 571-578.
- Knight, M., Ittiprasert, W., Odoemelam, E.C., Adema, C.M., Miller, A., Raghavan, N. *et al.* (2011). Non-random organization of the *Biomphalaria glabrata* genome in interphase Bge cells and the spatial repositioning of activated genes in cells co-cultured with *Schistosoma mansoni*. *International journal for parasitology*, 41, 61-70.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128, 693-705.
- Krauss, V. & Reuter, G. (2010). DNA methylation in *Drosophila*--a critical evaluation. *Progress in molecular biology and translational science*, 101, 177-191.
- Krautz-Peterson, G., Bhardwaj, R., Faghiri, Z., Tararam, C.A. & Skelly, P.J. (2010). RNA interference in schistosomes: machinery and methodology. In: *Parasitology*. Cambridge Univ Press, p. 485.
- Kriaucionis, S. & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 324, 929-930.
- Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, 319, 1827-1830.
- Kunert, N., Marhold, J., Stanke, J., Stach, D. & Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development*, 130, 5083-5090.
- Lachmann, M. & Jablonka, E. (1996). The inheritance of phenotypes: an adaptation to fluctuating environments. In: *J. theor. Biol.* Elsevier, pp. 1-9.
- Lamarck (1815). *Histoire Naturelle des Animaux sans Vertèbres. Histoire Naturelle des Animaux sans Vertèbres, Présentant les Caractères Généraux et Particuliers de ces Animaux, leur Distribution, leurs Classes, leurs Familles, leurs Genres, et la Citation des Principales Espèces qui s'y Rapportent... Volume, 5.*
- Lepesant, J.M., Roquis, D., Emans, R., Cosseau, C., Arancibia, N., Mitta, G. *et al.* (2012). Combination of de novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Experimental parasitology*, 130, 470-474.
- Lepesant, J.M.J., Grunau, C. & Cosseau, C. (2011). Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. In: *Memorias do Instituto Oswaldo Cruz* pp. 823-830.
- LeRoux, P.L. (1954). Hybridization of *Schistosoma mansoni* and *S. rodhaini*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 48, 3-4.
- Li, B., Carey, M. & Workman, J.L. (2007). The role of chromatin during transcription. *Cell*, 128, 707-719.

- Li, E., Bestor, T.H. & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69, 915-926.
- Lister, R. & Ecker, J.R. (2009). Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome research*, 19, 959-966.
- Lockyer, A., Olson, P., Østergaard, P., Rollinson, D., Johnston, D., Attwood, S. *et al.* (2003). The phylogeny of the Schistosomatidae based on three genes with emphasis on the interrelationships of *Schistosoma* Weinland, 1858. *Parasitology*, 126, 203-224.
- Loker, E.S. & Brant, S.V. (2006). Diversification, dioecy and dimorphism in schistosomes. *TRENDS in Parasitology*, 22, 521-528.
- Lovrecic, L., Maver, A., Zadel, M. & Peterlin, B. (2013). The Role of Epigenetics in Neurodegenerative Diseases.
- Luger, K. (2006). Dynamic nucleosomes. *Chromosome Research*, 14, 5-16.
- Lunt, D.H., Kumar, S., Koutsovoulos, G. & Blaxter, M.L. (2014). The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ*, 2, e356.
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C. & Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS biology*, 8, e1000506.
- Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Landmarks in Medical Genetics: Classic Papers with Commentaries*, 51, 226.
- Marhold, J., Rothe, N., Pauli, A., Mund, C., Kuehle, K., Brueckner, B. *et al.* (2004). Conservation of DNA methylation in dipteran insects. *Insect Mol. Biol.*, 13, 117-123.
- Mattick, J.S. & Makunin, I.V. (2006). Non-coding RNA. *Human molecular genetics*, 15, R17-R29.
- Maynard Smith, J. (1990). Models of a dual inheritance system. In: *Journal of theoretical biology*. Elsevier, pp. 41-53.
- Mayr, E. (1963). Animal species and evolution. *Animal species and their evolution*.
- Mehta, I.S., Amira, M., Harvey, A.J. & Bridger, J.M. (2010). Research Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G. *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448, 553-560.
- Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H. *et al.* (2004). The genome sequence of silkworm, *Bombyx mori*. *Dna Research*, 11, 27-35.

- Mitta, G., Adema, C.M., Gourbal, B., Loker, E.S. & Theron, A. (2012). Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Dev. Comp. Immunol.*, 37, 1-8.
- Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S. & Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, 303, 343-348.
- Mohamed, A.H., El-Din, A.T.S., Mohamed, A.M. & Habib, M.R. (2012). The relationship between genetic variability and the susceptibility of *Biomphalaria alexandrina* snails to *Schistosoma mansoni* infection. *Memórias do Instituto Oswaldo Cruz*, 107, 326-337.
- Moné, Y., Gourbal, B., Duval, D., Du Pasquier, L., Kieffer-Jaquinod, S. & Mitta, G. (2010). A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. In: *PLoS One*. Public Library of Science, p. e813.
- Moné, Y., Ribou, A.-C., Cosseau, C., Duval, D., Théron, A., Mitta, G. *et al.* (2011). An example of molecular co-evolution: Reactive oxygen species (ROS) and ROS scavenger levels in *Schistosoma mansoni*/*Biomphalaria glabrata* interactions. *International journal for parasitology*, 41, 721-730.
- Morgan, J., Dejong, R., Snyder, S., Mkoji, G. & Loker, E. (2001). *Schistosoma mansoni* and *Biomphalaria*: past history and future trends. *Parasitology*, 123, 211-228.
- Moullia, C., Le Brun, N., Loubes, C., Marin, R. & Renaud, F. (1995). Hybrid vigour against parasites in interspecific crosses between two mice species. *Heredity*, 74, 48-52.
- Mutani, A., Christensen, N.O. & Frandsen, F. (1985). A study of the biological characteristics of a hybrid line between male *Schistosoma haematobium* (Dar es Salaam, Tanzania) and female *S. intercalatum* (Edea, Cameroun). In: *Acta Tropica*, pp. 319-331.
- Nanty, L., Carbajosa, G., Heap, G.A., Ratnieks, F., van Heel, D.A., Down, T.A. *et al.* (2011). Comparative methylomics reveals gene-body H3K36me3 in *Drosophila* predicts DNA methylation and CpG landscapes in other invertebrates. In: *Genome Research*. Cold Spring Harbor Lab, pp. 1841-1850.
- Narlikar, G.J., Fan, H.-Y. & Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell*, 108, 475-487.
- Navarro, M., Peñate, X. & Landeira, D. (2007). Nuclear architecture underlying gene expression in *Trypanosoma brucei*. *Trends in microbiology*, 15, 263-270.
- Norouzitallab, P., Baruah, K., Vandegehuchte, M., Van Stappen, G., Catania, F., Bussche, J.V. *et al.* (2014). Environmental heat stress induces epigenetic transgenerational inheritance of robustness in parthenogenetic *Artemia* model. *The FASEB Journal*, fj. 14-252049.
- Nuruzzaman, M., Alam, M., Ahmed, M., Shohael, A., Biswas, M., Amin, M. *et al.* (2002). Studies on parental variability and heterosis in rice. *Pakistan J. Biol. Sci.*, 5, 1006-1009.

- Odoemelam, E.C. (2009). Genomic analysis of the fresh water mollusc *Biomphalaria glabrata* to understand host: parasite interactions. Brunel University.
- Okano, M., Bell, D.W., Haber, D.A. & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99, 247-257.
- Olins, A.L. & Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science*, 183, 330-332.
- Oliveira, K.C., Carvalho, M.L.P., Maracaja-Coutinho, V., Kitajima, J.P. & Verjovski-Almeida, S. (2011). Non-coding RNAs in schistosomes: an unexplored world. *Anais Da Academia Brasileira De Ciencias*, 83, 673-694.
- Pal, C. & Miklos, I. (1999). Epigenetic inheritance, genetic assimilation and speciation. In: *J. theor. Biol.* Elsevier, pp. 19-37.
- Pastor, W.A., Aravind, L. & Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Reviews Molecular Cell Biology*, 14, 341-356.
- Pedersen, A.G., Baldib, P., Chauvinb, Y. & Brunaka, S. (1999). The biology of eukaryotic promoter prediction - a review. In: *Computers & Chemistry*, p. 207.
- Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. In: *Biochemical Journal*. Portland Press Ltd, p. 781.
- Perez-Quintero, J., Bech Taberner, M. & Huerts Dionisio, L. (2004). Molluscs of the inland waters of the province of Huelva (SO Spain). *Iberus*, 22, 19-31.
- Perrin, C., Lepesant, J.M.J., Roger, E., Duval, D., Fneich, S., Thuillier, V. *et al.* (2013). *Schistosoma mansoni* Mucin Gene (SmPoMuc) Expression: Epigenetic Control to Shape Adaptation to a New Host. In: *PLOS Pathogens*. Public Library of Science, p. e1003571.
- Peserico, A. & Simone, C. (2010). Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *BioMed Research International*, 2011.
- Pitchford, R.J. (1961). Observations on a possible hybrid between the two schistosomes *S. haematobium* and *S. mattheei*. In: *Transactions of the Royal Society of tropical Medicine and Hygiene*. Elsevier, pp. 44-51.
- Planade, B., Lena, J.-P., Li, H., Plenet, S., Guegan, J.-F.o., Thomas, F. *et al.* (2009). Tracking a heterosis effect in the field: tadpole resistance to parasites in the water frog hybridogenetic complex. In: *Cambridge Univ Press*. Cambridge Univ Press, pp. 1003-1013.
- Polley, L., Hoberg, E. & Kutz, S. (2010). Climate change, parasites and shifting boundaries. *Acta Veterinaria Scandinavica*, 52, 0147-0152.
- Prell, H.H. & Day, P. (2001). *Plant-fungal pathogen interaction: a classical and molecular view*. Springer.

- Raddatz, G., Guzzardo, P.M., Olova, N., Fantappiè, M.R., Rampp, M., Schaefer, M. *et al.* (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences*, 110, 8627-8631.
- Raghavan, N. & Knight, M. (2006). The snail (*Biomphalaria glabrata*) genome project. *Trends in parasitology*, 22, 148-151.
- Raghavan, N., Tettelin, H., Miller, A., Hostetler, J., Tallon, L. & Knight, M. (2007). Nimbus (BgI): an active non-LTR retrotransposon of the *Schistosoma mansoni* snail host *Biomphalaria glabrata*. In: *International Journal for Parasitology*. Elsevier, pp. 1307-1318.
- Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I. & Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends in Genetics*, 18, 348-351.
- Ralph, S.A. & Scherf, A. (2005). The epigenetic control of antigenic variation in *Plasmodium falciparum*. *Current opinion in microbiology*, 8, 434-440.
- Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I. & Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*, 441, 469-474.
- Redmond, D.L., Knox, D.P., Newlands, G. & Smith, W.D. (1997). Molecular cloning and characterisation of a developmentally regulated putative metallopeptidase present in a host protective extract of *Haemonchus contortus*. *Molecular and biochemical parasitology*, 85, 77-87.
- Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K. & Bonner, W. (2002). Histone H2A variants h2ax and h2az. *Current opinion in genetics & development*, 12, 162-169.
- Richards, E.J. (2006). Inherited epigenetic variation - revisiting soft inheritance. In: *Nature Reviews Genetics*. Nature Publishing Group, pp. 395-401.
- Roberts, S.B. & Gavery, M.R. (2011). Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Frontiers in physiology*, 2.
- Roger, E., Gourbal, B., Grunau, C., Pierce, R.J., Galinier, R. & Mitta, G. (2008a). Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 157, 217-227.
- Roger, E., Grunau, C., Pierce, R.J., Hirai, H., Gourbal, B., Galinier, R. *et al.* (2008b). Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). In: *PLoS neglected tropical diseases*. Public Library of Science, p. e330.
- Roger, E., Mitta, G., Mone, Y., Bouchut, A., Rognon, A., Grunau, C. *et al.* (2008c). Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach. *Molecular and Biochemical Parasitology*, 157, 205-216.

- Rollinson, D. & Simpson, A.J. (1987). *The biology of schistosomes. From genes to latrines*. Academic Press Ltd.
- Roquis, D., Lepesant, J.M., Villafan, E., Boissier, J., Vieira, C., Cosseau, C. *et al.* (2014). Exposure to hycanthone alters chromatin structure around specific gene functions and specific repeats in *Schistosoma mansoni*. *Frontiers in genetics*, 5.
- Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S. *et al.* (2011). Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *The EMBO journal*, 30, 1928-1938.
- Sager, R. & Kitchin, R. (1975). Selective silencing of eukaryotic DNA. *Science*, 189, 426-433.
- Salzberg, A., Fisher, O., Siman-Tov, R. & Ankri, S. (2004). Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochemical and biophysical research communications*, 322, 465-469.
- Scherf, A., Lopez-Rubio, J.J. & Riviere, L. (2008). Antigenic variation in *Plasmodium falciparum*. *Annu. Rev. Microbiol.*, 62, 445-470.
- Schneeberger, K., Ossowski, S., Ott, F., Klein, J.D., Wang, X., Lanz, C. *et al.* (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proceedings of the National Academy of Sciences*, 108, 10249-10254.
- Scoville, A.G., Barnett, L.L., Bodbyl-Roels, S., Kelly, J.K. & Hileman, L.C. (2011). Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. *New Phytologist*, 191, 251-263.
- Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L. *et al.* (2012). Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *The Plant Cell Online*, 24, 875-892.
- Shull, G.H. (1914). Duplicate genes for capsule-form in *Bursa pastoris*. *Zeitschrift ind. Abst. u. Vererbungsl.*, 12, 97-149.
- Shull, G.H. (1948). What Is "Heterosis"? In: *Genetics*. Genetics Society of America, p. 439.
- Shull, G.H. & Gowen, J.W. (1952). Beginnings of the heterosis concept. In: *book* Iowa State College Press, pp. 14-48.
- Simpson, V.J., Johnson, T.E. & Hammen, R.F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic acids research*, 14, 6711-6719.
- Skelly, P.J., Da'dara, A. & Harn, D.A. (2003). Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *International Journal for Parasitology*, 33, 363-369.
- Smith, J.M. (1989). *Evolutionary genetics*. Oxford University Press.

- Sorensen, M.K., Norberg, E., Pedersen, J. & Christensen, L.G. (2008). Invited Review: Crossbreeding in Dairy Cattle: A Danish Perspective. In: *Journal of Dairy Science*. Elsevier, pp. 4116-4128.
- Southgate, V.R., Van Wijk, H.B. & Wright, C.A. (1976). Schistosomiasis at Loum, Cameroun; *Schistosoma haematobium*, *S. intercalatum* and their natural hybrid. In: *Zeitschrift für Parasitenkunde*. Springer, pp. 145-159.
- Srivastava, N., Gochhait, S., de Boer, P. & Bamezai, R.N. (2009). Role of H2AX in DNA damage response and human cancers. *Mutation Research/Reviews in Mutation Research*, 681, 180-188.
- Standen, O. (1953). The penetration of the cercariae of *Schistosoma mansoni* into the skin and lymphatics of the mouse. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 47, 292-298.
- Strahl, B.D. & Allis, C.D. (2000). The language of covalent histone modifications. *Nature*, 403, 41-45.
- Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B. *et al.* (2002). A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome research*, 12, 543-554.
- Sun, J., Wang, S., Li, C., Ren, Y. & Wang, J. (2014). Novel expression profiles of microRNAs suggest that specific miRNAs regulate gene expression for the sexual maturation of female *Schistosoma japonicum* after pairing. *Parasites & vectors*, 7, 177.
- Suzuki, M.M. & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9, 465-476.
- Suzuki, M.M., Kerr, A.R.W., De Sousa, D. & Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. In: *Genome Research*. Cold Spring Harbor Lab, pp. 625-631.
- Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D. & Schnable, P.S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. In: *PNAS*. National Acad Sciences, pp. 6805-6810.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y. *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324, 930-935.
- Tan, L. & Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*, 139, 1895-1902.
- Tanabe, H., Müller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M. *et al.* (2002). Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proceedings of the National Academy of Sciences*, 99, 4424-4429.
- Taylor, M.G. (1970). Hybridization experiments on five species of African schistosomes. In: *Journal of helminthology*. Cambridge Univ Press, pp. 253-314.

- Theodoropoulos, G., Hicks, S.J., Corfield, A.P., Miller, B.G. & Carrington, S.D. (2001). The role of mucins in host-parasite interactions: Part II-helminth parasites. *Trends in parasitology*, 17, 130-135.
- Theron, A. (1986). Chronobiology of schistosome development in the snail host. *Parasitology Today*, 2, 192-194.
- Théron, A. (1989). Hybrids between *Schistosoma mansoni* and *S. rodhaini*: characterization by cercarial emergence rhythms. In: *Journal of Parasitology*. Cambridge Univ Press, pp. 225-228.
- Théron, A. & Coustau, C. (2005). Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? In: *Journal of helminthology*. Cambridge Univ Press, pp. 187-192.
- Theron, A., Rognon, A., Gourbal, B. & Mitta, G. (2014). Multi-parasite host susceptibility and multi-host parasite infectivity: A new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni* compatibility polymorphism. *Infection, Genetics and Evolution*, 26, 80-88.
- Thomas, F., Lefèvre, T. & Raymond, M. (2010). *Biologie évolutive*.
- Thomas, H. & Gönner, R. (1977). The efficacy of praziquantel against cestodes in animals. *Zeitschrift für Parasitenkunde*, 52, 117-127.
- Tweedie, S., Charlton, J., Clark, V. & Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and cellular biology*, 17, 1469-1475.
- Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. (1982). The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS letters*, 146, 148-152.
- Valinluck, V. & Sowers, L.C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer research*, 67, 946-950.
- Van Der Biezen, E.A. & Jones, J.D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in biochemical sciences*, 23, 454-456.
- Van Valen, L. (1974). Molecular evolution as predicted by natural selection. *Journal of molecular evolution*, 3, 89-101.
- Verstrepen, K.J. & Fink, G.R. (2009). Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annual review of genetics*, 43, 1-24.
- Via, S. & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 505-522.
- Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150, 563-565.
- Waddington, C.H. (1953). Genetic assimilation of an acquired character. *Evolution*, 118-126.
- Waddington, C.H. (1957). *The strategy of the genes*. Routledge.

- Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D. *et al.* (2010). A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.*, 19, 215-228.
- Wang, Z., Schones, D.E. & Zhao, K. (2009). Characterization of human epigenomes. *Current opinion in genetics & development*, 19, 127-134.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S. *et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics*, 40, 897-903.
- Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M. *et al.* (2014). Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open biology*, 4, 140110.
- Woodruff, D.S. & Mulvey, M. (1997). Neotropical schistosomiasis: African affinities of the host snail *Biomphalaria glabrata* (Gastropoda: Planorbidae). *Biological Journal of the Linnean Society*, 60, 505-516.
- Wright, C.A. & Ross, G.C. (1980). Hybrids between *Schistosoma haematobium* and *S. matthei* and their identification by isoelectric focusing of enzymes. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene*. Elsevier, pp. 326-332.
- Wright, C.A., Southgate, V. Wijk, H.B. van, Moore, P. J. (1974). Hybrids between *Schistosoma haematobium* and *S. intercalatum* in Cameroun. *Trans. roy. Soc. trop. Med. Hyg.*, 68, 413-414.
- Wyatt, G. (1951). Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochem J.*
- Yoshida, M., Horinouchi, S. & Beppu, T. (1995). Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *Bioessays*, 17, 423-430.
- Yves, M. (2011). Bases moléculaires du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata*. In: *Academie de Montpellier Université de Perpignan*.
- Zemach, A., McDaniel, I.E., Silva, P. & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328, 916-919.
- Zerlotini, A., Heiges, M., Wang, H., Moraes, R.L., Dominitini, A.J., Ruiz, J.C. *et al.* (2009). SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic acids research*, 37, D579-D582.
- Zhang, S.M., Adema, C.M., Kepler, T.B. & Loker, E.S. (2004). Diversification of Ig superfamily genes in an invertebrate. *Science*, 305, 251-254.
- Zhang, Y. & Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & development*, 15, 2343-2360.
- Zhao, Y., Yu, S., Xing, C., Fan, S. & Song, M. (2008). Analysis of DNA methylation in cotton hybrids and their parents. *Molecular Biology*, 42, 169-178.

- Zhou, X.-N., Yang, G.-J., Yang, K., Wang, X.-H., Hong, Q.-B., Sun, L.-P. *et al.* (2008). Potential impact of climate change on schistosomiasis transmission in China. *The American journal of tropical medicine and hygiene*, 78, 188-194.
- Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. & Henikoff, S. (2006). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature genetics*, 39, 61-69.

Chapitre VII

Annexe

Chapitre VII : Annexe

Cosseau et ses collaborateurs ont mis au point les techniques permettant d'analyser les modifications des histones par immunoprécipitation de la chromatine (ChIP). Cette technique permet l'étude de la relation entre les modifications des histones et l'expression de certains gènes. Elle consiste en l'utilisation des **anticorps** qui reconnaissent spécifiquement des **marques d'histones**. Etant donné que les histones sont des protéines hautement conservées, un bon nombre d'anticorps commerciaux sont disponibles. Le protocole d'immunoprécipitation de la chromatine est disponible sur <http://methdb.univ-perp.fr/epievo/>.

Une étape de mise au point de la spécificité et de l'efficacité de l'anticorps sur le modèle étudié, est nécessaire. La spécificité de l'anticorps est déterminée au préalable par des expériences de Western blot, afin de vérifier que l'anticorps n'interagit qu'avec la protéine cible.

Dans le cas où un profil à plusieurs bandes est détecté par Western blot, cet anticorps est considéré comme aspécifique. D'autres anticorps reconnaissant la même cible seront testés par la suite. Le cas idéal est celui où nous détectons une bande unique à la taille attendue en KDa, dans ce cas l'anticorps est retenu. Même en l'absence d'un signal par Western blot, l'anticorps sera tout de même testé par Immunoprécipitation.

Lorsque l'anticorps est validé par Western blot, nous effectuons un ChIP pour la titration de l'anticorps (déterminer le ratio anticorps / chromatine). Pour ces expériences de titration nous appliquons à la même quantité du matériel biologique (de chromatine), des quantités croissantes de l'anticorps. L'anticorps doit être en excès pour que la quantité de chromatine immunoprécipitée ne dépende pas de la quantité d'anticorps utilisée.

Dans le cas d'un anticorps spécifique, la quantité d'ADN immunoprécipité devient constante à un moment donné, même si la quantité de l'anticorps augmente. Cependant, lorsqu'un anticorps est aspécifique, la quantité d'ADN immunoprécipité continue à augmenter avec l'augmentation de la quantité d'anticorps.

Seuls les bons anticorps validés en Western blot et en ChIP sont retenus pour les expérimentations. Les étapes de la mise au point de l'utilisation des anticorps sont couteuses en anticorps, en matériel biologique et en temps, mais une fois l'anticorps est validé, les immunoprécipitations sont reproductibles.

Au cours de la thèse, j'ai effectué des mises au point de certains anticorps sur notre modèle, incluant les trois espèces *S. mansoni*, *B. glabrata* et *M. auratus*.

J'ai eu l'occasion aussi de mettre au point la spécificité de différents anticorps par Western blot, sur 2 autres modèles différents.

(i) Par collaboration avec **Klaus Brehm** de *Institut für Hygiene und Mikrobiologie*, j'ai testé 15 anticorps (dont l'histone H3, et différentes marques d'histones telles que H3K9me3, H3K4me3, H3K36me2 etc.) sur le cestode *Echinococcus multilocularis*.

(ii) Par collaboration avec **Laetitia Perfus-Barbeoch** de l'Université de Nice Sophia-Antipolis, j'ai testé 12 anticorps sur le nématode *Meloidogyne incognita*.

Pour cette dernière collaboration (ii), le but du travail était de caractériser les bases moléculaires de l'information épigénétique chez ce nématode. J'ai contribué à la détection des histones et des marques d'histones par Western blot, utilisant différents anticorps. Ceci a fait l'objet d'une publication dans **Frontiers in PHYSIOLOGY** (Publication attachée, figure 2).

Publication:

Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode *Meloidogyne incognita*

Laetitia Perfus-Barbeoch Zurletto, Philippe Castagnone-Sereno, Michael Reichelt, Sara Fneich, David Roquis, Loris Pratx, Celine Cosseau, Christoph Grunau and Pierre Abad

Frontiers in Invertebrate Physiology 06 Jun 2014





Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode *Meloidogyne incognita*

Laetitia Perfus-Barbeoch^{1,2,3*}, Philippe Castagnone-Sereno^{1,2,3}, Michael Reichelt⁴, Sara Fneich^{5,6}, David Roquis^{5,6}, Loris Pratz^{1,2,3}, Céline Cosseau^{5,6}, Christoph Grunau^{5,6} and Pierre Abad^{1,2,3}

¹ INRA, Institut Sophia Agrobiotech, UMR 1355 ISA, Sophia-Antipolis, France

² CNRS, Institut Sophia Agrobiotech, UMR 7254 ISA, Sophia-Antipolis, France

³ Institut Sophia Agrobiotech, Université de Nice Sophia-Antipolis, UMR ISA, Sophia-Antipolis, France

⁴ Abteilung Biochemie, MPI für Chemische Ökologie, Jena, Germany

⁵ Ecologie et Evolution des Interactions, Université de Perpignan Via Domitia, Perpignan, France

⁶ Ecologie et Evolution des Interactions, CNRS, UMR5244, Perpignan, France

Edited by:

Mackenzie Gaverty, University of Washington, USA

Reviewed by:

Guilherme Corrêa De Oliveira,

Fundação Oswaldo

Cruz - FIOCRUZ-Minas, Brazil

Elena Gómez-Díaz, Emory

University, USA

*Correspondence:

Laetitia Perfus-Barbeoch, INRA, UMR 1355 ISA, CNRS, UMR 7254 ISA, Institut Sophia Agrobiotech, Université de Nice Sophia-Antipolis, 400 route des Chappes, BP167, Sophia Antipolis cedex, Alpes-Maritimes, France
e-mail: zurletto@sophia.inra.fr

Root-knot nematodes of the genus *Meloidogyne* are biotrophic plant parasites that exhibit different life cycles and reproduction modes, ranging from classical amphimixis to obligatory mitotic parthenogenesis (apomixis), depending on the species. *Meloidogyne incognita*, an apomictic species, exhibits a worldwide distribution and a wide host range affecting more than 3000 plant species. Furthermore, evidences suggest that apomixis does not prevent *M. incognita* from adapting to its environment in contrast to what is expected from mitotic parthenogenesis that should theoretically produce clonal progenies. This raises questions about mechanisms of genome plasticity leading to genetic variation and adaptive evolution in apomictic animals. We reasoned that epigenetic mechanisms might in part be responsible for the generation of phenotypic variants that provide potential for rapid adaptation. We established therefore a pipeline to investigate the principal carriers of epigenetic information, DNA methylation and post-translational histone modifications. Even if *M. incognita* possesses the epigenetic machinery i.e., chromatin modifying enzymes, 5-methyl-cytosine and 5-hydroxy-methyl-cytosine content is absent or very weak. In contrast, we demonstrated that the canonical histone modifications are present and chromatin shows typical nucleosome structure. This work is the first characterization of carriers of epigenetic information in *M. incognita* and constitutes a preamble to further investigate if *M. incognita* development and its adaptation to plant hosts are under epigenetic control. Our pipeline should allow performing similar types of studies in any non-model organism.

Keywords: epigenetics, *Meloidogyne incognita*, chromatin, DNA-methylation, histone modification

INTRODUCTION

Epigenetics is the study of heritable changes in gene expression and function that cannot be explained by changes in DNA sequence. These molecular mechanisms can be influenced by the environment and the individual history as well as being potentially transferable from a generation to the next, with a reversible character (Youngson and Whitelaw, 2008). Organisms are forced to adapt to the heterogeneity of the environment and the phenotypic plasticity constitutes a major adaptation to compulsory constraints. Since the system of reproduction of an organism conditions its evolutionary dynamics as well as the demography and the spatial genetic structuring of its populations, it is of particular relevance to study epigenetic variation in asexually reproducing organisms. Indeed, in spite of a genetic homogeneity in asexually reproducing organisms, an important phenotypic variability can be observed between individuals of the same clonal lineage found in different environments. The influence of the epigenetic mechanisms could bring an explanation to

this paradox (reviewed in Verhoeven and Preite, 2014). Similarly, epigenetic modifications may provide an accessory source of fast-acting, reversible, and readily available phenotypic variations that can shape host-pathogen interactions (reviewed in Gómez-Díaz et al., 2012). Despite growing evidences on the role of epigenetic phenomena on a wide range of biological phenomena, many questions remain for the involvement of epigenetics in plasticity and adaptation.

Root-knot nematodes (RKN) of the genus *Meloidogyne* are obligatory plant parasites that constitute major agricultural pests worldwide. They infect almost all cultivated plants and establish an intimate interaction with their hosts inducing the re-differentiation of root cells into hypertrophied and multinucleate feeding cells (Caillaud et al., 2008). So far, the most efficient control measure against RKN was the use of chemical substances called nematicides. However, because of their toxic side effects on the environment, nematicides have recently been banned from use. Novel approaches to control them are sorely needed. Among

RKN species, some are strict or facultative amphimictic while others reproduce exclusively by mitotic parthenogenesis (apomixis) (Castagnone-Sereno, 2006). Our model, *Meloidogyne incognita*, reproduces in an asexual way by parthenogenesis without meiosis. Genetically identical individuals develop from *M. incognita* females and form virtually clonal populations. Although these clones share the same genetic heritage, modifications of their phenotype can be observed when they are exposed to unfavorable environments (Castagnone-Sereno et al., 1994). These phenotypes can be completely different such as juveniles molting into either female or male, but also show for instance a switch of virulence leading to infection of resistant hosts. A first step toward a better understanding of plant-nematode interaction was the sequencing of the genome of *M. incognita*. Analysis of this reference genome highlighted characteristics, which could be at the origin of the parasitic success of this species (Abad et al., 2008; Danchin et al., 2010).

Epigenetics changes refer to a set of molecular processes that can affect gene expression such as regulatory processes mediated by remodeling of chromatin structure through methylation of cytosine residues in the DNA and chemical modifications of histone proteins, in particular their acetylation or methylation (Jablonka and Lamb, 2002). On one hand, DNA methylation refers to a chemical modification of genomic DNA by the addition of a methyl (-CH₃) group to specific nucleotide bases. The most common form of DNA methylation is cytosine methylation, occurring predominantly in CpGs in animal genomes (Suzuki and Bird, 2008). However, because of lack of *de novo* methylation machinery, DNA methylation process is unlikely to occur in nematodes, including *C. elegans* but with the exception of *Trichinella spiralis* (Simpson et al., 1986; Wenzel et al., 2011; Gao et al., 2012, 2014). On the other hand, post-translational modifications of histone amino acid side chains (PTMs) are known to mediate epigenetic regulation of gene expression and development in animals (Li et al., 2007; Luger et al., 2012; Turner, 2012). For instance, PTMs, such as acetylation of histone H3 at lysine 9, are typically found in euchromatin and often associated with active transcription, whereas other modifications, such as methylation of lysine 9 or lysine 27 of histone H3, are found in heterochromatin and related to gene repression (Jenuwein and Allis, 2001). One major role for epigenetic variation in evolution would be to promote phenotypic variability, and allow populations to widely explore new environmental conditions. This could lead to a rapid adaptation to environmental changes or colonization of new environments. The interactions between parasites and their host are models of choice to study these mechanisms because the selective pressures are strong and the evolution is fast. *M. incognita* constitutes an ideal model to study these mechanisms, especially as clones can be obtained following the naturally occurring asexual multiplication in host plants. Whereas variation at the genetic level has been studied in *M. incognita*, less is known about the extent and function of epigenetic variation. Previous analyses on multiple *M. incognita* populations have shown that genomic polymorphism could not be directly correlated to observed phenotypes (Semblat et al., 2000).

Many non-model organisms of ecological and economical importance such as *M. incognita* suffer from a lack of even

rudimentary knowledge about their epigenetic information carriers. To address this problem, we describe here a straightforward 2 × 3-step procedure to perform a fundamental analysis of elements of the chromatin marking system. We first performed a whole genome bioinformatics search for chromatin-modifying enzymes such as those involved in DNA methylation, we evaluate the degree of DNA methylation by calculating the CpG observed/expected ratio in *M. incognita* transcriptome and by measuring DNA methylation levels of 5-methylcytosine (5-mC) and 5-hydroxy methylcytosine (5-hmC) in genomic DNA by LC-MS/MS. We then used Western blots to verify the presence of histone and histone modifications and tested critical parameters of a standardized N-ChIP procedure that need to be adapted for each model.

ANALYSIS OF DNA METHYLATION

DNA methylation is a common feature of many but not all genomes. We recommend starting with this feature.

STEP 1: *IN-SILICO* SEARCHES OF CHROMATIN-REMODELING ENZYMES

To understand whether *M. incognita* possesses the capacity to methylate DNA, we first conducted reciprocal BLASTp searches combined to OrthoMCL analysis to identify proteins that might be related to known chromatin-remodeling enzyme orthologs (Table 1). Significant hits were defined as those satisfying the following criteria: *E*-value < 10⁻⁵ and the aligned segments covering at least 30% of the sequence length of the hit. OrthoMCL has been applied to the proteome data set from *M. incognita* and three publicly available nematode genomes (*C. elegans*, *T. spiralis*, and *Pristionchus pacificus*). Animal DNA methyltransferases can be subdivided into three subfamilies, DNMT-1, DNMT-2 and DNMT-3, based on sequence similarity (Kumar et al., 1994; Hendrich and Tweedie, 2003). DNMT-1 has been regarded as a “maintenance” methyltransferase, whereas DNMT-3 functions as “*de novo*” methyltransferase and some evidence suggest that DNMT-2 enzymes are active cytosine-5'-methyltransferases (Kunert et al., 2003). Species exhibiting functional DNA methylation generally encode a complete set of DNMTs in their genomes whereas species lacking DNA methylation, such as *C. elegans*, have lost DNA methylation enzymes from their genomes (Gutierrez and Sommer, 2004; Suzuki and Bird, 2008). BLASTp search of the *M. incognita* whole proteome (Supplementary Material) was performed with DNMT-1, DNMT-2, DNMT-3 protein sequences from nematodes (Table 1): the *Trichinella spiralis* DNMT-1 (accession number EFV58204.1); *P. pacificus* Ppa-DNMT-2 (accession number AY766101.1) and *T. spiralis* DNMT-22 (accession number EFV60295.1); *T. spiralis* DNMT-3 (accession number EFV54759.1). As for the 11 other nematodes in which DNA methylation has been studied, with exception of *T. spiralis* (Gao et al., 2012), *M. incognita* did not possess *de novo* methylation machinery, DNMT-3 nor DNMT-2. However, our data revealed the existence of a potential DNMT-1 orthologous protein annotated in *M. incognita*, as well as methyl-CpG binding domain (MBDs) proteins, which encodes another essential component of the methylation system (Table 1). We found that Minc01117 could be ortholog to the maintenance methyltransferase DNMT-1 from other species. Sites of DNA methylation are occupied

Table 1 | *M. incognita* putative orthologs for methyltransferases (DNMTs) and methyl-CpG binding domain (MBDs) proteins.

Protein	Species	Accession number	<i>M. incognita</i> putative orthologs
DNA METHYLTRANSFERASES (DNMTs)			
DNMT-1	<i>Trichinella spiralis</i>	EFV58204.1	Minc01117
DNMT-2	<i>Trichinella spiralis</i>	EFV60295.1	<i>M. incognita</i> absent from OrthoMCL group
	<i>Pristionchus pacificus</i>	AY766101.1	
DNMT-3	<i>Trichinella spiralis</i>	EFV54759.1	<i>M. incognita</i> absent from OrthoMCL group
METHYL-CpG BINDING DOMAIN 2 (MBDs)			
Cel-MBD-2	<i>Caenorhabditis elegans</i>	C27A12.10	Minc14299 Minc14778
Cbr-MBD-2	<i>Caenorhabditis briggsae</i>	CBP11474	
Ppa-MBD-2	<i>Pristionchus pacificus</i>	AY766102.1	

by various proteins, including MBD proteins, which recruit the enzymatic machinery to establish silent chromatin. Nematodes such as *C. elegans* (Cel-MBD-2, accession number C27A12.10), *C. briggsae* (Cbr-MBD-2, accession number CBP11474) and *P. pacificus* (Ppa-MBD-2, accession number AY766102.1) all contain an mbd-2-like gene (Gutierrez and Sommer, 2004). We found that two predicted *M. incognita* proteins, Minc14299 and Minc14778, could be orthologs to the MBD-2 like protein.

STEP 2: CpG OBSERVED/EXPECTED RATIO (CpGo/e) ANALYSIS

To evaluate the degree of DNA methylation, we calculated the empirical distribution of CpGo/e ratios for *M. incognita* transcriptome. A total of 63,838 expressed sequenced tags (EST) of *M. incognita* that were downloaded from NCBI-dbEST public database (Supplementary Material), filtered by size (≥ 500 bp) leading to 41,649 sequences of which 5000 were randomly chosen for CpG observed/expected ratio (CpGo/e) analysis. Most of the data came from *M. incognita* strain Morelos maintained at INRA Sophia Antipolis and recently published (Abad et al., 2008; Jaouannet et al., 2012). To validate the robustness of CpGo/e ratio method, the same experiment was conducted on *C. elegans* (initially 8722 ESTs), the model nematode lacking DNA methylation process, *D. melanogaster* (38,110 ESTs) and on *A. mellifera* (10,157 ESTs). *D. melanogaster* has been previously described for its lack of DNA methylation at late stages of development (Urieli-Shoval et al., 1982; Lyko et al., 2000) while DNA methylation occurs in *A. mellifera* (Elango et al., 2009). The CpGo/e ratio was used as a proxy to estimate the intragenic DNA methylation content as previously described in *Biomphalaria glabrata* (Fneich et al., 2013). In *C. elegans*, the CpGo/e ratio profile exhibited an unskewed Gaussian distribution with an estimated mean value close to 1 (0.93 ± 0.26) as expected for species lacking such process (Figure 1A). A similar result was obtained with

D. melanogaster late stage ESTs (Estimated mean value 0.92 ± 0.17 ; Figure 1B). In contrast, the CpGo/e ratio in *A. mellifera* transcripts can be approximated by a mixture of two Gaussian distributions (Figure 1C). This pattern has also been described in the mollusc *B. glabrata* where transcripts are divided in low and high methylated genes (Fneich et al., 2013). *M. incognita* CpGo/e ratio profile did not perfectly fit *C. elegans* profile (Figure 1D). In *M. incognita*, the CpGo/e ratio profile also exhibited a skewed near-Gaussian distribution but with a slight shift on the left compared to *C. elegans* profile. The estimated mean value for *M. incognita* transcriptome was inferior to 1 (0.77 ± 0.25). However, careful analysis of the curve fitting suggests the presence of two populations of sequences: one with a mean value around 0.9 and another with a mean value around 0.7. These results could indicate the presence of two populations of genes in *M. incognita*: one with no methylation and one with very low methylation.

STEP 3: GLOBAL DNA METHYLATION ANALYSIS BY LC-MS/MS

Genomic DNA used for liquid chromatography-mass spectrometry analysis was prepared from *M. incognita* eggs with the CTAB method (Winnepeninckx et al., 1993) and stored at -80°C until use. The levels of 5-methylcytosine (5-mC) and 5-hydroxy methylcytosine (5-hmC) were determined by liquid chromatography-mass spectrometry LC-MS/MS analysis as their deoxyribonucleosides as described in (Fneich et al., 2013). Neither 5-methyl-cytosine nor 5-hydrox-methyl-cytosine could be detected.

Taken together, our results show that DNA methylation is absent or very low in *M. incognita*.

ANALYSIS OF NUCLEOSOMES, HISTONES, AND HISTONE MODIFICATIONS

In the absence of DNA methylation, PTMs could represent one of the primary epigenetic transcriptional control mechanisms. Such a process has been shown to play a key role in the regulation of gene transcription of parasites, e.g., the apicomplexan family of protozoa and the metazoan *Schistosoma* (Mourão et al., 2012; Perrin et al., 2013). The importance of these modifications in regulating diverse developmental programs in parasites motivated us to study their importance in *M. incognita* in the framework of this study. Experiments were performed on *M. incognita* strain “Morelos” which have been previously used for the whole genome sequencing of this nematode (Abad et al., 2008). Because *M. incognita* is an obligatory plant parasite, it has to be maintained on plants to complete its life cycle. We used tomatoes, *Solanum esculentum* cultivar Saint Pierre, grown at 20°C in a greenhouse as *M. incognita* host for this study. Moreover, samples that we collected are considered to be clonal because of the mitotic parthenogenetic mode of reproduction of *M. incognita* and since the lab population was originated from the progeny of a single female. To grow and produce nematodes, one-month-old tomato plants are inoculated with *M. incognita* second-stage juveniles (J2s). As previously described by Rosso et al., 1999, eggs were collected from 7 week-old infected tomato roots after grinding, sterilizing (0.5% NaOCl) and filtering steps. Extracted eggs were kept either to hatch J2s or to be aliquoted and stored at -80°C for further experiments (e.g., genomic DNA extraction

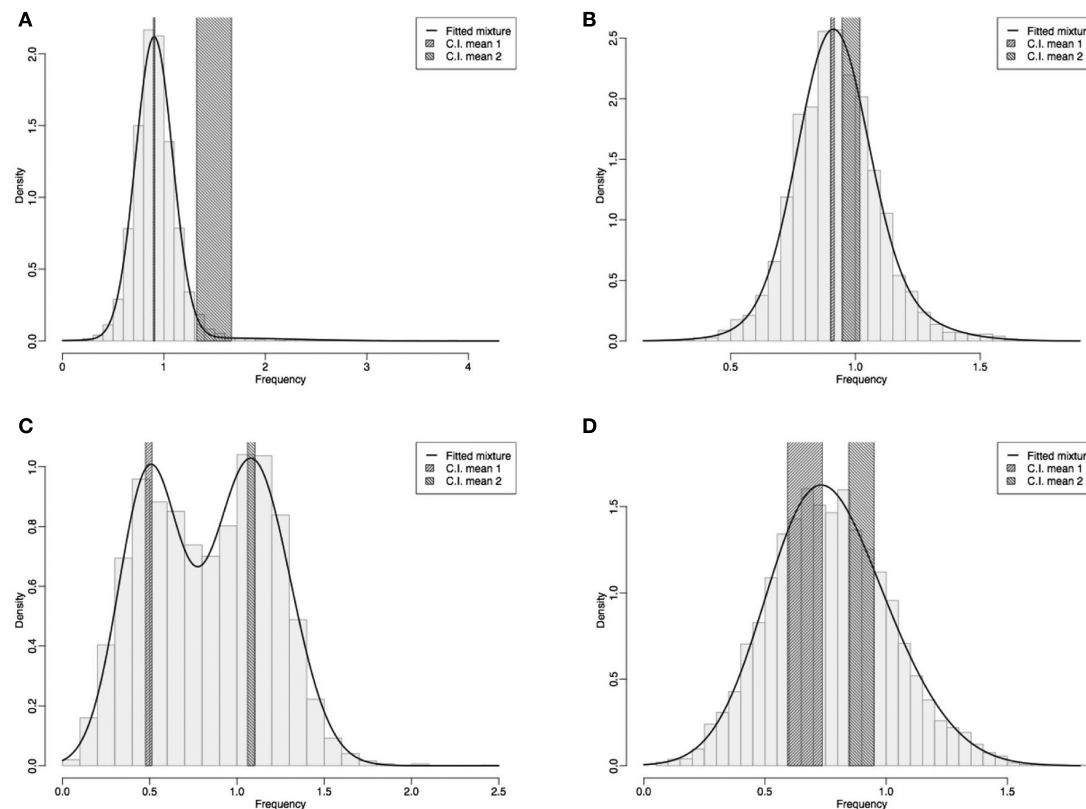


FIGURE 1 | Frequency distribution of CpG observed/expected ratio (CpGo/e) in four different species. (A) *C. elegans* and (B) *D. melanogaster*, and on (C) *A. melifera* and (D) *M. incognita*. CpGo/e ratio was measured as a proxy to estimate the CpG methylation in transcripts from EST data. X axis:

CpGo/e ratio, Y-axis density (frequency distribution) of ESTs. The figure displays a histogram of Bg GUA CpGo/e ratios with a fitted mixture distribution. The gray shaded bars represent 95% confidence intervals for the two mean values.

and chromatin immuno-precipitation). To produce J2s, extracted eggs were kept on 10 μ m sieve in autoclaved tap water, aerated with an air pump, at room temperature. After 5 days, J2s hatched and were collected by centrifugation (13000 g, 1 min). J2 samples were stored at -80°C for further experiments (e.g., Western blot and chromatin immuno-precipitation).

STEP 1: DETECTION OF CANONICAL HISTONES AND HISTONE MODIFICATIONS BY WESTERN BLOTS

The nucleosome is subject to a dizzying array of posttranslational modifications, which work alone or in combination to constitute a histone code that regulates chromatin structure and function (Jenuwein and Allis, 2001). The four core histones, H2A, H2B, H3, and H4, which are responsible for folding DNA into nucleosomes, have been described as very highly conserved throughout evolution (Mariño-Ramírez et al., 2006). This high conservation allowed us to test commercial antibodies that react with mammals but also with a wide range of species. In some species, chromatin is condensed around the centromere and characterized by several methyl groups on lysine amino acid (K) at 9 and 20 position of histone H3 and H4 respectively. These marks are noted H3K9Me2/3 and H4K20Me3 and can be recognized by specific

antibodies. Before immuno-precipitation of *M. incognita* DNA, we tested 12 commercial antibodies, directed against histone H3 and several H3 and H4 modifications, on *M. incognita* J2s with western blots.

Western blots were performed as previously described by Azzi et al. (2009). Briefly, about one thousand *M. incognita* J2s were re-suspended in denaturation buffer, treated by sonication and boiled 5 min at 99°C . Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Amersham RPN203D) by the semi-dry method (SEMI-PHOR Bio-Rad). The membrane was blocked overnight at 4°C in blocking buffer and incubated with one of the 12 following commercial antibodies (1 $\mu\text{g}/\mu\text{l}$): Anti-histone H3 (H3, abcam ab1791 and Active Motif 39164); Anti-Histone H3 dimethylated at lysine 4 (H3K4Me2, Abcam ab32356) or at lysine 36 (H3K36Me2, Abcam ab9049); Anti-Histone H3 trimethylated at lysine 4 (H3K4Me3, Millipore 04-745 and Abcam ab8580), at lysine 9 (H3K9Me3, Abcam ab8898 and Upstate) or at lysine 27 (H3K27Me3, Diagenode pAB-069-050); Anti-Histone H3 acetylated at lysine 9 (H3K9Ac, Upstate); Anti-Histone H4 trimethylated at lysine 20 (H4K20Me3, abcam ab9053) and Anti-hyperacetylated Histone H4 (H4PentaAc, Upstate). Bands were

revealed by Enhanced Chemical Luminescence (ECL Pierce) and direct exposure to X-ray film (Amersham).

The antibodies should give a unique staining on the membrane (one unique band) to be selected for DNA immunoprecipitation. Three antibodies, Anti-H3K4Me3 (Abcam), Anti-H3K9Me3 (Abcam and Upstate), did not give any signal (**Supplementary Figure 2A**). One antibody, Anti-H4PentaAc gave a non-specific signal with many high molecular weight bands (**Supplementary Figure 2B**). Two antibodies, Anti-H3K27Me3 and Anti-H3K36Me2 gave two distinct bands (**Supplementary Figure 2C**). Six antibodies, Anti-histone H3 (Abcam and Active), Anti-H3K4Me2, Anti-H3K4Me3 (Millipore), Anti-H3K9Ac and Anti-H4K20Me3 gave a unique signal at 17 kDa (**Figure 2**) as expected for histones (Bártová et al., 2005).

We conclude that canonical histones and histone modifications are present and that six commercially available antibodies could be used for immuno-precipitation of *M. incognita* chromatin.

STEP 2: CONTROLLED FRAGMENTATION INTO NUCLEOSOMES

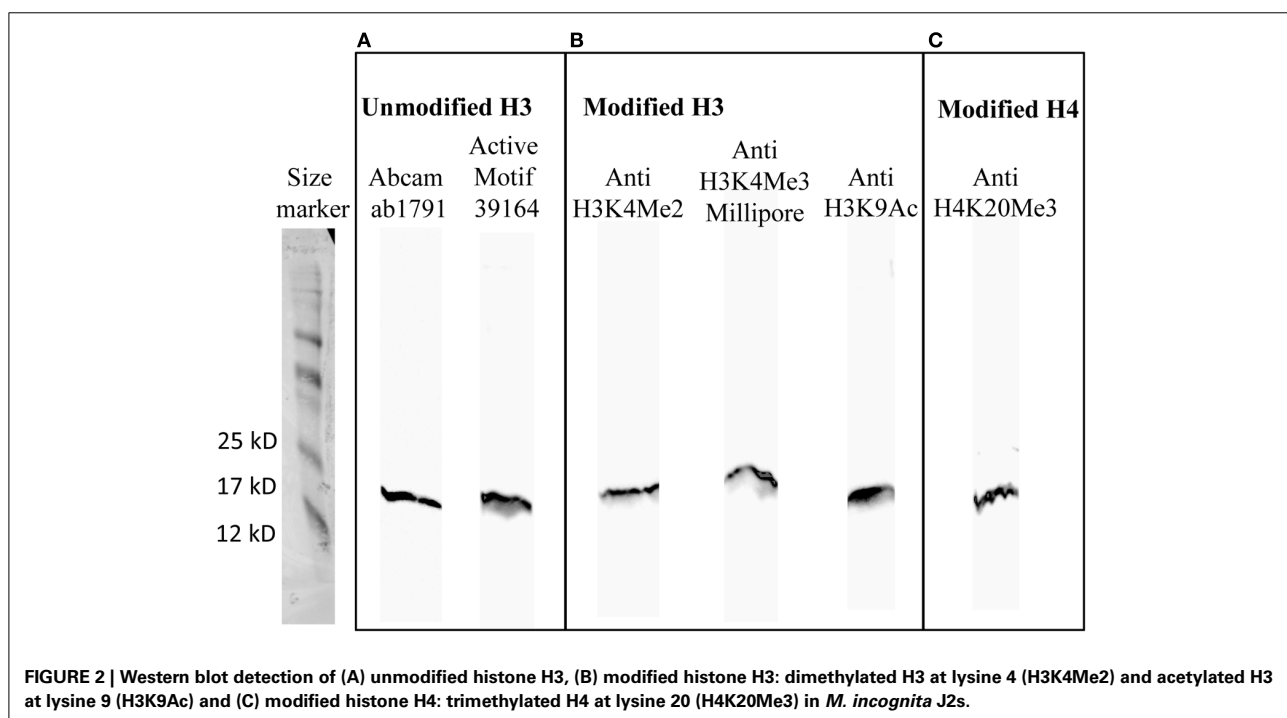
Because little is known about chromatin organization in *M. incognita*, we started by testing partial Micrococcal Nuclease (MNase) digestion of chromatin to reveal its nucleosome structure, which is the fundamental unit of chromatin. Because DNA portions of nucleosome core particles are less accessible for MNase than linking sections, DNA gets digested into fragments of lengths equal to multiplicity of distance between mono-nucleosomes. First, to access nuclei, we tested different grinding times, from 3 to 10 min, for cell lysis in Dounce homogenizer. As a measure for grinding efficiency, we colored nuclei using Hoechst, a

fluorescent stain that binds strongly DNA. Also, because *M. incognita* life-stages are really different in shape and composition, we tested these grinding conditions on both J2s and eggs. Best results were obtained with 10 min total in Dounce homogenizer, distributed as 5 min grinding following by 5 min at rest (**Figure 3A**). Then, we tested different MNase concentrations (from 15 to 30 U; **Figure 3B** and **Supplementary Figure 3**) and incubation times (0, 2, 4, 5, 6, and 7 min). Typical mono-nucleosome structure was observed with fragments about 150 bp and multiplicity of this size (**Figure 3B**). Finally, 5 min incubation with 1 μ l MNase (15 U) was chosen as the standard conditions to perform further steps of immuno-precipitation of the chromatin in *M. incognita*.

STEP 3: CHROMATIN IMMUNOPRECIPITATION (ChIP)

To immunoprecipitate *M. incognita* DNA, we used the ChIP procedure adapted for non-model organisms by Cosseau et al. (2009). Each step of the procedure (**Supplementary Material**) had to be customized for *M. incognita*. Basically, we tested different grinding times for cell lysis and different Micrococcal nuclease (MNase, Affymetrix, USB 70196Y) concentrations and incubation times (see Results). Nuclei were stained with Hoechst 33342 fluorescent stain (1/2000 dilution; Invitrogen H1399) and observed under microscope (40X, Nikon T1-SM) before and after grinding.

Because the antibody should be in excess over the histone of interest for ChIP, we tested protein/chromatin precipitation efficiency in *M. incognita* with increasing antibody volume ranging from 0 to 16 μ l (1 μ g/ μ l), for each of 3 tested antibodies (among the 6 antibodies that could be used): Anti-H3 (Abcam), Anti-H3K4Me3 (Millipore), and Anti-H4K20Me3. For each antibody concentration, DNA from the bound and unbound fractions



was extracted with phenol/chloroform so we could quantify the sequences associated with each antibody by qPCR. In other species, histone marks such as H3K4Me3 were enriched around the transcription start sites (TSSs) of genes and were associated with active transcription (Du et al., 2013). For this reason, primers were designed to match genomic region close to TSSs of

M. incognita housekeeping genes, Actin and GAPDH (Minc06773 and Minc12412, respectively; **Table 2**). Quantitative Real-time PCR (qPCR) analyses were performed using the LightCycler 2.0 system (Roche Applied Science) and LightCycler Faststart DNA Master SYBR Green I kit (Roche Applied Science). qPCR amplification was done in a final volume of 10 μ l composed of 2.5 μ l of immuno-precipitated chromatin, 1.5 μ l H₂O, 0.5 μ l of each primer 10 μ M and 5 μ l of master mix. The following Light-Cycler run protocol was used: denaturation, 95°C, 10 min; amplification and quantification (repeated 40 times), 95°C for 10 s, 62°C for 5 s, 72°C for 16 s; melting curve, 60–95°C with a heating rate of 0.11°C/s and continuous fluorescence measurement, and a cooling step to 40°C. For each reaction, the cycle threshold (Ct) was determined using the “2nd derivative” method of the LightCycler 480 Software release 1.5. Primers were designed for this experiment with the Primer3 web-interface (**Table 2**). Quality and specificity controls were performed for each qPCR product (**Supplementary Figure 1**): amplification of a unique 150 bp band was verified by electrophoresis separation through a 2% agarose gel; amplicons were sequenced to verify that they matched expected sequence (GATC-biotech; **Table 2**); Primer efficiency was calculated based on Ct value for increasing genomic DNA concentration (from 0 to 1 ng/ μ l); Melting curve and negative first derivative of the melting-curve were recorded for each qPCR product to validate that only one product was amplified (one peak was observed).

The amount of target DNA recovered in the immuno-precipitated fraction was quantified by calculating the percent input recovery (% IR) normalized with the percent input recovery obtained with the housekeeping gene as previously described by Cosseau et al. (2009).

Also, because of previous MNase digestion, the amplicon size has to be no longer than 150 pb to match *M. incognita* mono-nucleosome length. With *M. incognita* DNA, antibodies Anti-H3 and Anti-H4K20Me3 never reached saturation up to 16 μ l (**Figure 4**). This could mean that these antibodies are not specific to histone H3 and histone H4 trimethylated at lysine 20 but recognize other proteins/marks. So, the more antibody was added, the more proteins interacted which generated a background signal detected thanks to qPCR. On the contrary, the Anti-H3K4Me3 antibody successfully immuno-precipitated *M.*

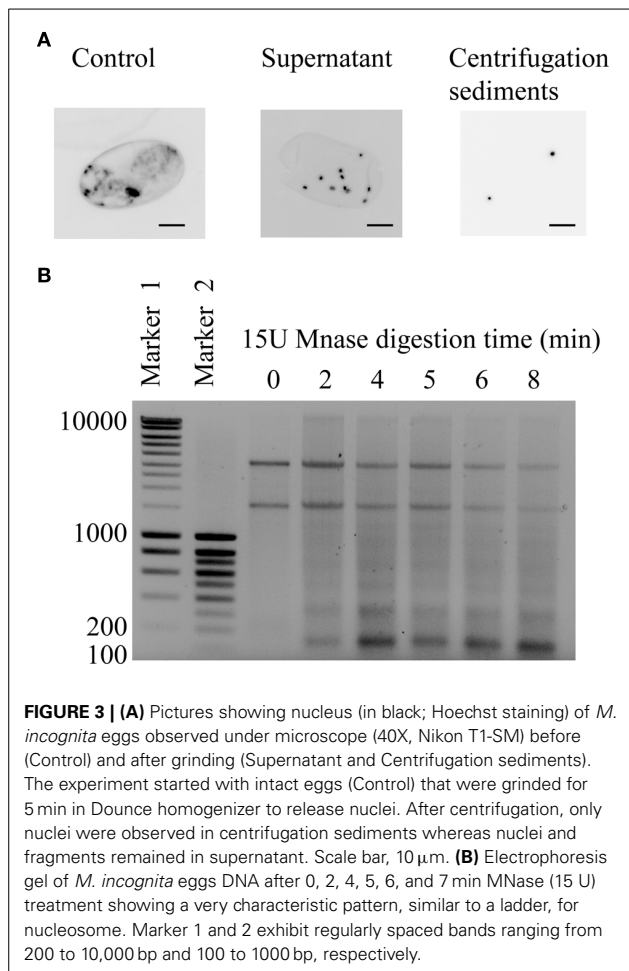
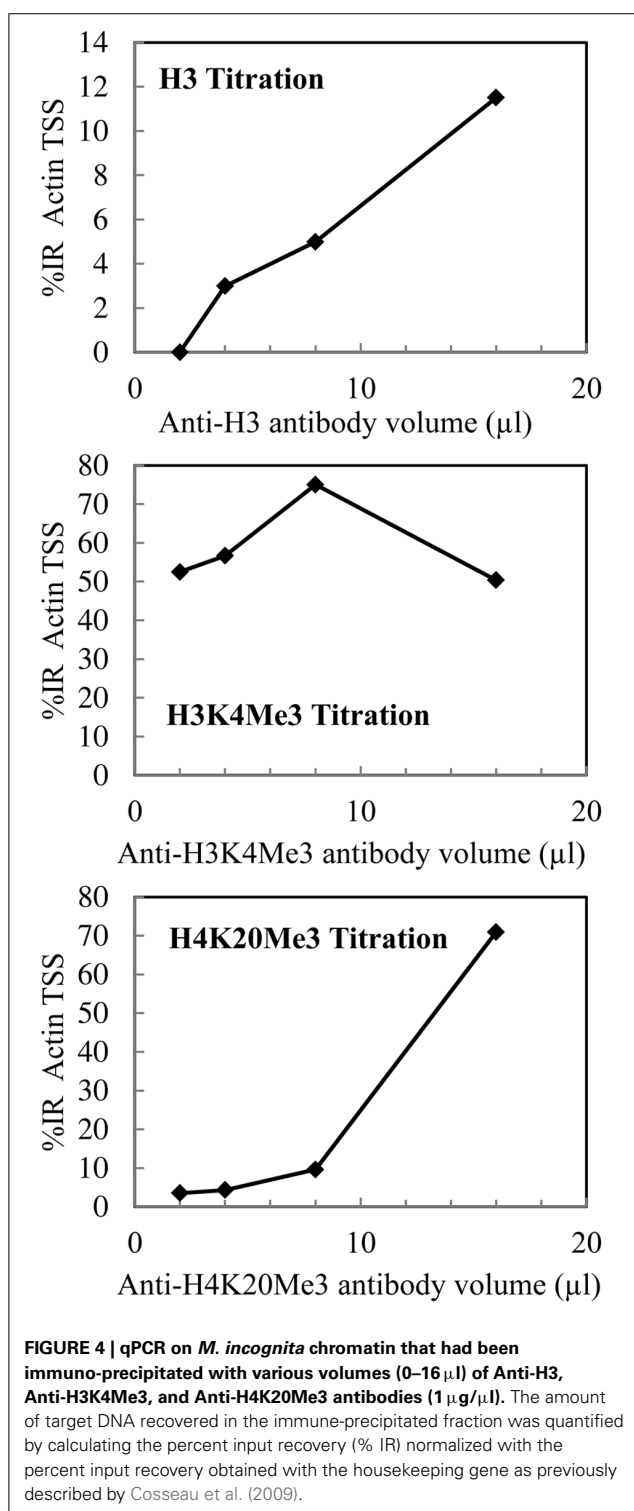


Table 2 | Amplicon sequences and primer sequences designed for qPCR.

Name	Sequence	Length (bp)	Product size (bp)
<i>M. incognita</i> GAPDH amplicon	CGTGCAGCGGTTGAGAAGGATACTGTCCAAGTTGTTGCTGTCAATGACCCGTTTCATTGATCTTGACTATATGGTTTGGGGGAGACTTTCTATTAATAACTCCAATAACTTTTAGGTCTATATGTTAACTATGATCCACCCACGGACGC		
Mi_ChIP_GAPDH_F	CGTGCAGCGGTTGAGAAGGA	20	150
Mi_ChIP_GAPDH_R	GCGTCCGTGGGTGGAATCAT	20	
<i>M. incognita</i> Actin amplicon	AAGATGGATGAAGAGGTAGCCGCCCTTGTGTGGATAATGGCAGTGGAATGTGCAAGGTTAGTTAAAAATGCCCTTTTTCGATTAAAGTTCTGTTGTTTTTAAAGGCTGGATTGCGGGGTGATGATGCTCCTCGTGCCGTTTTTCCA		
Mi_ChIP_Actin_F	AAGATGGATGAAGAGGTAGCCGCC	25	150
Mi_ChIP_Actin_R	TGGAAAAACGGCACGAGGAGCA	22	



incognita chromatin and reached saturation at 8 μl (Figure 4). Titration saturation for higher antibody concentrations indicates that the antibody is specific to the histone mark (Cosseau et al., 2009). Therefore, this antibody could specifically

immuno-precipitate *M. incognita* and could be used for further analysis.

To summarize, we optimized the ChIP protocol for *M. incognita*, based on the previously published method for non-model organisms (Cosseau et al., 2009). First, we have optimized lysis and extraction of nuclei for *M. incognita* to 10 min grinding in digestion buffer. Then we found that the optimum time for MNase digestion (15U), to obtain fragmented chromatin into mono-nucleosome to penta-nucleosomes, a size range that is optimal for ChIP, was 5 min. Finally, qPCR experiments indicated that N-ChIP procedure should be performed with an excess of 8 μl antibody Anti-H3K4Me3. Now that chromatin immuno-precipitation has been optimized in *M. incognita* it opens new perspectives to further investigate the genome-wide distribution of PTMs by high-throughput sequencing (N-ChIP-Seq).

DISCUSSION

In the past few years, genome projects have demonstrated that DNA methylation is far more widespread than one would expect from the lack of DNA methylation in non-mammalian model organisms, such as the model nematode *C. elegans*. Moreover, functional DNA methylation has been recently reported in the parasitic nematode *T. spiralis* (Gao et al., 2012, 2014). Altogether, this makes us wonder what DNA methylation picture in *M. incognita* could be.

The *de novo* DNA methylation process can produce new methylated sites without pre-existing pattern by affecting cytosine nucleotide (Goll and Bestor, 2005). Depending on CG density, some of these *de novo* sites can then be maintained by DNMT maintenance methyltransferases. CpG under representation is generally seen as sign of DNA methylation (Bird, 1980). CpG o/e suggests that DNA methylation was present as a subset of genes in *M. incognita* ancestral species, that these sequences were obtained by horizontal gene transfer from a species with DNA methylation or that the EST database was contaminated. In the first case, *M. incognita* might have lost DNA methylation enzymes, which may have been present in the ancestral nematode, from its genome, as what is thought for *C. elegans* (Gutierrez and Sommer, 2004; Suzuki and Bird, 2008). Nevertheless, the observed CpG bias could also be due to other mechanisms (Jabbari and Bernardi, 2004). By LC-MS DNA methylation could not be detected in *M. incognita* that goes in line with the absence of gene coding for *de novo* DNA methylation activity.

The nucleosome is one of the most highly conserved structures known in eukaryotes, which is consistent with the particle's fundamental role in packaging DNA into the nucleus (Turner, 1993, 2005). After prolonged digestion, the MNase degrades the chromatin to its fundamental repeating unit, called the nucleosome. In *M. incognita*, DNA fragment observed micrococcal nuclease digestion of 150 bp define the nucleosome that is consistent with what is known of the "primary structure" of chromatin with DNA fragment of length around the ~145–150 pb (Kornberg and Thomas, 1974; Luger et al., 1997; Rando, 2011). However, it is becoming clear that chromatin structures are not nearly as uniform and regular as previously assumed: nucleosomes vary in their histone protein components due to the incorporation of variant histones and PTMs (Luger et al., 2012;

Turner, 2012). All PTMs, reversible for the most part, either change nucleosome structure directly by affecting histone proteins and DNA interactions or indirectly by recruiting binding proteins that act on the underlying chromatin structure, as has been proposed in the “histone code” hypothesis (Strahl and Allis, 2000). However, in spite of the high conservation of histone proteins along evolution, their degree of modification varies enormously according to the species, the developmental stage, the tissue and the cellular cycle phase (Strahl and Allis, 2000; Trojer and Reinberg, 2007; Luger et al., 2012; Turner, 2012). Here, we provided evidences for a “canonical nucleosome structure” in *M. incognita* with recognition of both unmodified and modified histone H3 as well as modified histone H4. Both euchromatin and heterochromatin regions contain specific PTMs and binding proteins. For instance, in *C. elegans*, heterochromatin is enriched in trimethylated H3 histones at lysine 56 (H3K56Me3) and lysine 9 (H3K9Me3; Jack et al., 2013). Ultimately, to identify features of transcriptional networks that regulate developmental processes, ChIP-sequencing was completed genome-wide in *C. elegans* (Niu et al., 2011). We expect to build on this foundation, together with this preliminary study, for further evaluation of transcriptional regulatory mechanisms during *M. incognita* development.

In addition, it is to be noted that nematodes could be divided in monocentric species and holocentric species (Subirana and Messegue, 2013). *C. elegans* as well as *M. incognita* belong to the latter in which the centromeric function spreads out over the whole chromosomes (Triantaphyllou, 1983; Dernburg, 2001; Melters et al., 2012). Moreover, changes in chromosome number and structural rearrangements have been reported in *Meloidogyne* species (Triantaphyllou, 1981; Castagnone-Sereno, 2006; Abad et al., 2008). Being able to study position and histone protein components of *M. incognita* nucleosomes can be used to better understand diverse biological processes including DNA replication, recombination, mutation and repair. Because of the intimate relationship between nucleosome locations and these processes, we anticipate that ChIP-seq experiments in *M. incognita* will shed more light on genome evolution that seems to be crucial for an organism whose mode of reproduction is obligatory mitotic parthenogenesis (apomixis).

One major role for epigenetic variation in evolution would be to promote phenotypic variability, and allow populations to widely explore new environmental conditions. However, the detailed mechanisms thanks to which parasites evade host immunity for instance are poorly understood. *M. incognita* constitutes an ideal model to study these mechanisms, especially as inheritance of the nematode virulence does not segregate in accordance with Mendel's laws together with asexual multiplication in host plants (Castagnone-Sereno et al., 1994). Because clones can be obtained, genetic variability is reduced. As a result, *M. incognita* provides support for the study of epigenetic heredity distinct from the canonical Mendelian rules. Recently, it became apparent that parasitism success is partly accomplished by epigenetic means in several parasites, including, among others, protozoans *Plasmodium* and trematodes *Schistosoma*. More especially, PTMs plays a key role in the regulation of gene transcription of these parasites (Mourão et al., 2012; Perrin et al., 2013). For

instance, it has been conclusively shown that PTMs, especially acetylations and methylations, and the propagation of heterochromatin away from the telomeres control the transcriptional switching between VAR genes that will trigger antigenic variation and enable immune evasion by the parasite, *Plasmodium falciparum* (Freitas-Junior et al., 2005; Chookajorn et al., 2007; Malmquist et al., 2012; Jiang et al., 2013). In *Schistosoma mansoni*, N-ChIP-seq experiments have been successfully performed to compare chromatin structure at different stages of the parasite life cycle, on different strains and on different sexes (Caby and Pierce, 2009; Cosseau et al., 2009; Lepesant et al., 2011, 2012). These advances in understanding *S. mansoni* epigenetic regulation have lead to the identification of promising new targets for the development of new drug to combat schistosomiasis (Pierce et al., 2012; Lancelot et al., 2013; Marek et al., 2013).

The importance of PTMs in regulating crucial developmental processes in parasites combined with the recent resolution of the genome sequences of *M. incognita* make the study of epigenetic machinery possible and susceptible to break new ground.

CONCLUSION

For the first time in *M. incognita*, this study details DNA methylation and nucleosome structure, carriers of epigenetic information. Genomic DNA of *M. incognita* is not methylated. The general characteristics of nucleosome structure that we observed in the *M. incognita* genome are in accordance to what epigenetic studies have reported from other invertebrate species. Histone modifications, important markers of function and chromatin state, were identified in *M. incognita*. This study opens the way for analyzing epigenetic mechanisms in *M. incognita* at a whole-genome scale to identify new biological processes involved in the generation of phenotypic variation in absence of sexual reproduction in an ecological and economical important model.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Céline Cosseau, Christoph Grunau, Laetitia Perfus-Barbeoch, Pierre Abad, Philippe Castagnone-Sereno. Performed the experiments: David Roquis, Laetitia Perfus-Barbeoch, Loris Pratz, Michael Reichelt, Sara Fneich. Analyzed the data: Céline Cosseau, Christoph Grunau, Laetitia Perfus-Barbeoch, Pierre Abad, Philippe Castagnone-Sereno. Wrote the paper: Céline Cosseau, Christoph Grunau, David Roquis, Laetitia Perfus-Barbeoch, Michael Reichelt, Pierre Abad, Philippe Castagnone-Sereno, Sara Fneich.

ACKNOWLEDGMENTS

The authors want to thank Nathalie Marteu (INRA, Sophia Antipolis, France) for growing plant and nematode materials, Etienne G. J. Danchin (INRA, Sophia Antipolis, France) for help with OrthoMCL analysis and Bruno Favery (INRA, Sophia Antipolis, France) for critical reading of the manuscript. This work was supported by the French Government (National Research Agency, ANR) through the “Investments for the Future” LABEX SIGNALIFE: program reference # ANR-11-LABX-0028-0.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fphys.2014.00211/abstract>

Supplementary material : Links to public databases.

Expressed sequenced tags (EST) of *M. incognita* from NCBI-dBEST: <http://www.ncbi.nlm.nih.gov/nucest/?term=meloidogyne+incognita>

Primer3 web-interface: www.bioinformatics.nl/primer3plus/

M. incognita whole proteome: http://www6.inra.fr/meloidogyne_incognita/Genomic-resources/Blast

Supplementary Figure 1 | Quality and specificity controls for qPCR analysis.

(A) Electrophoresis gel of 150 bp genomic DNA amplified with primers specifically designed for ChIP validation. Markers 1 and 2 exhibit regularly spaced bands ranging from 200 to 10,000 bp and 100 to 1000 bp, respectively. **(B)** Primer efficiency was assessed on increasing genomic DNA concentration (from 0 to 1 ng/μl): increase in fluorescence intensity is proportional to the increase in amplicon concentration. **(C)** Melting curves and negative first derivative of the melting-curve: presence of a single homogeneous melt peak confirms specific amplification.

Supplementary Figure 2 | Western blots showing non-specific signals in *M. incognita* J2s.

(A) Anti-H3K4Me3 or Anti-H3K9Me3 antibodies. **(B)** Anti-H4PentaAc antibody. **(C)** Anti-H3K27Me3 or Anti-H3K36Me2 antibodies.

Supplementary Figure 3 | Electrophoresis gel of *M. incognita* eggs DNA after 0, 2, 4, 5, 6, and 7 min MNase (30U).

Marker 1 and 2 exhibit regularly spaced bands ranging from 200 to 10,000 bp and 100 to 1000 bp, respectively.

REFERENCES

- Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., et al. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* 26, 909–915. doi: 10.1038/nbt.1482
- Azzi, A., Cosseau, C., and Grunau, C. (2009). *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp. Parasitol.* 121, 288–291. doi: 10.1016/j.exppara.2008.11.010
- Bártová, E., Pacherník, J., Harnicarová, A., Kovarik, A., Kovariková, M., Hofmanová, J., et al. (2005). Nuclear levels and patterns of histone H3 modification and HP1 proteins after inhibition of histone deacetylases. *J. Cell Sci.* 118, 5035–5046. doi: 10.1242/jcs.02621
- Bird, A. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8, 1499–1504. doi: 10.1093/nar/8.7.1499
- Caby, S., and Pierce, R. J. (2009). Quantitative chromatin immunoprecipitation (Q-ChIP) applied to *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 166, 77–80. doi: 10.1016/j.molbiopara.2009.02.014
- Caillaud, M. C., Dubreuil, G., Quentin, M., Perfus-Barbeoch, L., Lecomte, P., de Almeida Engler, J., et al. (2008). Root-knot nematodes manipulate plant cell functions during a compatible interaction. *J. Plant Physiol.* 165, 104–113. doi: 10.1016/j.jplph.2007.05.007
- Castagnone-Sereno, P. (2006). Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity* 96, 282–289. doi: 10.1038/sj.hdy.6800794
- Castagnone-Sereno, P., Wajnberg, E., Bongiovanni, M., Leroy, F., and Dalmasso, A. (1994). Genetic variation in *Meloidogyne incognita* virulence against the tomato Mi resistance gene: evidence from isofemale line selection studies. *Theor. Appl. Genet.* 88, 749–753. doi: 10.1007/BF01253980
- Chookajorn, T., Dzikowski, R., Frank, M., Li, F., Jiwan, A. Z., Hartl, D. L., et al. (2007). Epigenetic memory at malaria virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 899–902. doi: 10.1073/pnas.0609084103
- Cosseau, C., Azzi, A., Smith, K., Freitag, M., Mitta, G., and Grunau, C. (2009). Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Mol. Biochem. Parasitol.* 166, 70–76. doi: 10.1016/j.molbiopara.2009.02.015
- Danchin, E. G., Rosso, M. N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., et al. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17651–17656. doi: 10.1073/pnas.1008486107
- Dernburg, A. F. (2001). Here, there, and everywhere: kinetochore function on holocentric chromosomes. *J. Cell Biol.* 153, F33–F38. doi: 10.1083/jcb.153.6.F33
- Du, Z., Li, H., Wei, Q., Zhao, X., Wang, C., Zhu, Q., et al. (2013). Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. Japonica. *Mol. Plant.* 6, 1463–1472. doi: 10.1093/mp/sst018
- Elango, N., Hunt, B. G., Goodisman, M. A., and Yi, S. V. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11206–11211. doi: 10.1073/pnas.0900301106
- Fneich, S., Dheilly, N., Adema, C., Rognon, A., Reichelt, M., Bulla, J., et al. (2013). 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*. *Parasit. Vectors* 6:167. doi: 10.1186/1756-3305-6-167
- Freitas-Junior, L. H., Hernandez-Rivas, R., Ralph, S. A., Montiel-Condado, D., Ruvalcaba-Salazar, O. K., Rojas-Meza, A. P., et al. (2005). Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell.* 121, 25–36. doi: 10.1016/j.cell.2005.01.037
- Gao, F., Liu, X., Wu, X. P., Wang, X. L., Gong, D., Lu, H., et al. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biol.* 13, R100. doi: 10.1186/gb-2012-13-10-r100
- Gao, F., Wang, R., and Liu, M. (2014). *Trichinella spiralis*, potential model nematode for epigenetics and its implication in metazoan parasitism. *Front. Physiol.* 4:410. doi: 10.3389/fphys.2013.00410
- Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721
- Gómez-Díaz, E., Jordá, M., Peinado, M. A., and Rivero, A. (2012). Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS Pathog.* 8:e1003007. doi: 10.1371/journal.ppat.1003007
- Gutierrez, A., and Sommer, R. J. (2004). Evolution of *dnmt-2* and *mbd-2*-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res.* 32, 6388–6396. doi: 10.1093/nar/gkh982
- Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* 19, 269–277. doi: 10.1016/S0168-9525(03)00080-5
- Jabbari, K., and Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* 333, 143–149. doi: 10.1016/j.gene.2004.02.043
- Jablonka, E., and Lamb, M. J. (2002). The changing concept of epigenetics. *Ann. N. Y. Acad. Sci.* 981, 82–96. doi: 10.1111/j.1749-6632.2002.tb04913.x
- Jack, A. P., Bussemer, S., Hahn, M., Pünzeler, S., Snyder, M., Wells, M., et al. (2013). H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. *PLoS ONE* 8:e51765. doi: 10.1371/journal.pone.0051765
- Jaouannet, M., Perfus-Barbeoch, L., Deleury, E., Magliano, M., Engler, G., Vieira, P., et al. (2012). A root-knot nematode-secreted protein is injected into giant cells and targeted to the nuclei. *New Phytol.* 194, 924–931. doi: 10.1111/j.1469-8137.2012.04164.x
- Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–1080. doi: 10.1126/science.1063127
- Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., et al. (2013). PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* 499, 223–227. doi: 10.1038/nature12361
- Kornberg, R. D., and Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science* 184, 865–868. doi: 10.1126/science.184.4139.865
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., et al. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 22, 1–10. doi: 10.1093/nar/22.1.1
- Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A *Dnmt2*-like protein mediates DNA methylation in *Drosophila*. *Development* 130, 5083–5090. doi: 10.1242/dev.00716
- Lancelot, J., Caby, S., Dubois-Abdesselem, F., Vanderstraete, M., Troleat, J., Oliveira, G., et al. (2013). *Schistosoma mansoni* Sirtuins: characterization and

- potential as chemotherapeutic targets. *PLoS Negl. Trop. Dis.* 7:e2428. doi: 10.1371/journal.pntd.0002428
- Lepesant, J. M., Cosseau, C., Boissier, J., Freitag, M., Portela, J., Climent, D., et al. (2012). Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biol.* 13, R14. doi: 10.1186/gb-2012-13-2-r14
- Lepesant, J. M., Grunau, C., and Cosseau, C. (2011). Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. *Mem. Inst. Oswaldo Cruz.* 106, 823–830. doi: 10.1590/S0074-02762012000100022
- Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719. doi: 10.1016/j.cell.2007.01.015
- Luger, K., Dechassa, M. L., and Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* 13, 436–447. doi: 10.1038/nrm3382
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. doi: 10.1038/38444
- Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540. doi: 10.1038/35046205
- Malmquist, N. A., Moss, T. A., Mecheri, S., Scherf, A., and Fuchter, M. J. (2012). Small-molecule histone methyltransferase inhibitors display rapid antimalarial activity against all blood stage forms in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16708–16713. doi: 10.1073/pnas.1205414109
- Marek, M., Kannan, S., Hauser, A. T., Moraes Mourão, M., Caby, S., Cura, V., et al. (2013). Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog.* 9:e1003645. doi: 10.1371/journal.ppat.1003645
- Mariño-Ramírez, L., Jordan, I. K., and Landsman, D. (2006). Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biol.* 7:R122. doi: 10.1186/gb-2006-7-12-r122
- Melters, D. P., Palui, L. V., Korf, I. F., and Chan, S. W. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* 20, 579–593. doi: 10.1007/s10577-012-9292-1
- Mourão, M. M., Grunau, C., LoVerde, P. T., Jones, M. K., and Oliveira, G. (2012). Recent advances in *Schistosoma* genomics. *Parasite Immunol.* 34, 151–162. doi: 10.1111/j.1365-3024.2011.01349.x
- Niu, W., Lu, Z. J., Zhong, M., Sarov, M., Murray, J. L., Brdlik, C. M., et al. (2011). Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*. *Genome Res.* 21, 245–254. doi: 10.1101/gr.114587.110
- Perrin, C., Lepesant, J. M., Roger, E., Duval, D., Fneich, S., Thuillier, V., et al. (2013). *Schistosoma mansoni* mucin gene (SmPoMuc) expression: epigenetic control to shape adaptation to a new host. *PLoS Pathog.* 9:e1003571. doi: 10.1371/journal.ppat.1003571
- Pierce, R. J., Dubois-Abdesselem, F., Lancelot, J., Andrade, L., and Oliveira, G. (2012). Targeting schistosome histone modifying enzymes for drug development. *Curr. Pharm. Des.* 18, 3567–3578. doi: 10.2174/138161212801327248
- Rando, O. J. (2011). Genome-wide measurement of histone H3 replacement dynamics in yeast. *Methods Mol. Biol.* 759, 41–60. doi: 10.1007/978-1-61779-173-4_3
- Rosso, M. N., Favery, B., Pottie, C., Arthaud, L., De Boer, J. M., Hussey, R. S., et al. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Mol. Plant-Microbe Interact.* 12, 585–591. doi: 10.1094/MPMI.1999.12.7.585
- Semblat, J. P., Bongiovanni, M., Wajnberg, E., Dalmasso, A., Abad, P., and Castagnone-Sereno, P. (2000). Virulence and molecular diversity of parthenogenetic root-knot nematodes, *Meloidogyne* spp. *Heredity* 84, 81–89. doi: 10.1046/j.1365-2540.2000.00633.x
- Simpson, V. J., Johnson, T. E., and Hammen, R. F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* 14, 6711–6719. doi: 10.1093/nar/14.16.6711
- Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. doi: 10.1038/47412
- Subirana, J. A., and Messegue, X. (2013). A satellite explosion in the genome of holocentric nematodes. *PLoS ONE* 8:e62221. doi: 10.1371/journal.pone.0062221
- Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465–476. doi: 10.1038/nrg2341
- Triantaphyllou, A. C. (1981). Oogenesis and the Chromosomes of the Parthenogenetic Root-knot Nematode *Meloidogyne incognita*. *J. Nematol.* 13, 95–104.
- Triantaphyllou, A. C. (1983). “Cytogenetic aspects of nematode evolution,” in *Concepts in Nematode Systematics*, eds A. R. Stone, H. M. Platt, and L. F. Khalil (London: Academic Press), 55–71.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13. doi: 10.1016/j.molcel.2007.09.011
- Turner, B. M. (1993). Decoding the nucleosome. *Cell* 75, 5–8. doi: 10.1016/0092-8674(93)90673-E
- Turner, B. M. (2005). Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat. Struct. Mol. Biol.* 12, 110–112. doi: 10.1038/nsmb0205-110
- Turner, B. M. (2012). The adjustable nucleosome: an epigenetic signaling module. *Trends Genet.* 28, 436–444. doi: 10.1016/j.tig.2012.04.003
- Urieli-Shoval, S., Gruenbaum, Y., Sedat, J., and Razin, A. (1982). The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett.* 146, 148–152. doi: 10.1016/0014-5793(82)80723-0
- Verhoeven, K. J., and Preite, V. (2014). Epigenetic variation in asexually reproducing organisms. *Evolution* 68, 644–655. doi: 10.1111/evo.12320
- Wenzel, D., Palladino, F., and Jedrusik-Bode, M. (2011). Epigenetics in *C. elegans*: facts and challenges. *Genesis* 49, 647–661. doi: 10.1002/dvg.20762
- Winnepeninckx, B., Backeljau, T., and De Wachter, R. (1993). Extraction of high molecular weight DNA from molluscs. *Trends Genet.* 9, 407. doi: 10.1016/0168-9525(93)90102-N
- Youngson, N. A., and Whitelaw, E. (2008). Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.* 9, 233–257. doi: 10.1146/annurev.genom.9.081307.164445

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 December 2013; accepted: 15 May 2014; published online: 06 June 2014.
Citation: Perfus-Barbeoch L, Castagnone-Sereno P, Reichelt M, Fneich S, Roquis D, Pratz L, Cosseau C, Grunau C and Abad P (2014) Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode *Meloidogyne incognita*. *Front. Physiol.* 5:211. doi: 10.3389/fphys.2014.00211
This article was submitted to *Invertebrate Physiology*, a section of the journal *Frontiers in Physiology*.

Copyright © 2014 Perfus-Barbeoch, Castagnone-Sereno, Reichelt, Fneich, Roquis, Pratz, Cosseau, Grunau and Abad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Résumé :

La variabilité phénotypique est définie comme la capacité d'une espèce donnée à produire des variants phénotypiques à partir d'un seul génotype, sous l'influence de l'environnement. Certains de ces variants sont héréditaires. L'origine de la variabilité phénotypique a toujours constitué un grand débat parmi les scientifiques. Jusqu'à récemment, il était communément accepté que la diversité génétique soit la seule source de variabilité phénotypique. Cependant, les études récentes montrent que des variations épigénétiques pourraient être une source alternative pour les variants phénotypiques, sans modifier la séquence de l'ADN. Dans les interactions hôte / parasite, les parasites exercent des pressions sélectives sur leurs hôtes et *vice versa*, conduisant à une véritable course aux armements entre les deux partenaires. Une telle interaction nécessite une adaptation rapide où chaque partenaire doit présenter la capacité d'exprimer de nouveaux variants phénotypiques. Nous proposons que des variations épigénétiques puissent jouer un rôle important dans la genèse de la variabilité phénotypique. *Schistosoma mansoni* est un parasite humain responsable de la bilharziose intestinale. Au cours de la thèse, nous nous sommes intéressés à l'interaction du parasite *S. mansoni* avec son hôte intermédiaire le mollusque *Biomphalaria glabrata*. Les deux principaux objectifs de cette thèse ont été de : (i) Déterminer le poids relatif de l'épigénétique et de la génétique dans l'expression des variants phénotypiques chez le parasite. (ii) Initier les travaux de recherche sur les mécanismes épigénétiques existant chez le mollusque hôte.

Les résultats de nos travaux ont montré que les structures chromatiniques via des modifications post-traductionnelles des histones contribuent à modifier la variabilité phénotypique chez *S. mansoni*. Par ailleurs, nous avons montré que les variants phénotypiques générés confèrent au parasite une meilleure fitness. Cette dernière a été traduite par l'augmentation de sa compatibilité vis-à-vis de *B. glabrata*. Finalement, l'étude de l'hérédité des modifications épigénétiques a montré une transmission non mendélienne. En ce qui concerne *B. glabrata*, nous avons été les premiers à mettre en évidence la méthylation de l'ADN chez ce mollusque. Au niveau de son génome, 2% des cytosines totales sont méthylées.

Mots clés : variants phénotypiques, épigénétique, *Schistosoma mansoni*, *Biomphalaria glabrata*, méthylation de l'ADN, compatibilité, *SmPoMucs*, hérédité.

Abstract:

Phenotypic variability is defined as the capacity of a given population to produce phenotypic variants under the influence of the environment. Some of the phenotypic variants are heritable. It is generally assumed that the genetic variations are the sole source of heritable phenotypic variants. However, recent studies have shown that epigenetic variations can provide alternative source for phenotypic variants without change in DNA sequence. In host / parasite interactions, parasites impose selective pressures on their hosts and *vice versa*, leading to a genuine arms race between both partners. Such interaction leads to rapid adaptation where each partner has to evolve the capacity to express new phenotypic variants. We propose that epigenetic variations play an important role in the genesis of phenotypic variability. *Schistosoma mansoni* is a human parasite that causes intestinal schistosomiasis. During the PhD project, I was interested in the interaction of *S. mansoni* with its intermediate host *Biomphalaria glabrata*. The two main goals of this PhD project were: (i) To determine the relative weight of epigenetics and genetics in the expression of phenotypic variants in the parasite. (ii) To initiate the investigation of epigenetic mechanisms in the host.

The results show that histone modifications *i.e.* changes in the epigenetic information are indeed a source of phenotypic variants in *S. mansoni*. These phenotypic variants confer a higher fitness to the parasite, by increasing its compatibility towards the mollusk. Finally, studying the heritability of epigenetic changes showed a non-mendelian segregation. For *B. glabrata*, I was the first to show the DNA methylation in the snail. 2% of total cytosines are methylated in his genome.

Keywords: phenotypic variants, epigenetics, *Schistosoma mansoni*, *Biomphalaria glabrata*, DNA methylation, compatibility, *SmPoMucs*, heritability.



Laboratoire Ecologie et Evolution des
Interactions
UMR 5244 CNRS-UPVD
52 Avenue Paul Alduy 66860 Perpignan
<http://2ei.univ-perp.fr/>



UPVD
Université de Perpignan Via Domitia